

UNIVERSIDAD AUTÓNOMA DE MADRID
ESCUELA POLITÉCNICA SUPERIOR



Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación

TRABAJO FIN DE GRADO

IDENTIFICACIÓN DE MATERIALES UTILIZANDO EL SENSOR KINECT

Miguel Basarte Mena.
Tutor: Marcos Escudero Viñolo.
Ponente: Jesús Bescós Cano.

Junio 2017

IDENTIFICACIÓN DE MATERIALES UTILIZANDO EL SENSOR KINECT

Miguel Basarte Mena.
Tutor: Marcos Escudero Viñolo.
Ponente: Jesús Bescós Cano.



Video Processing and Understanding Lab
Departamento de Tecnología Electrónica y de las Comunicaciones
Escuela Politécnica Superior
Universidad Autónoma de Madrid
Junio 2017

Trabajo parcialmente financiado por el Ministerio de Economía y Competitividad del Gobierno de España bajo el proyecto TEC2014-53176-R (HAVideo) (2015-2017)



Resumen

En este trabajo de fin de grado se busca conseguir utilizar información frecuencial en una secuencia de imágenes para lograr identificar materiales de forma automática. Además, también se busca estudiar el impacto que puede tener añadir imágenes con información de luz infrarroja y profundidad (aparte de las imágenes RGB) en el desarrollo de esta tarea mencionada.

Para ello, se comienza creando un nuevo *dataset* con ayuda del sensor *Kinect*, compuesto por imágenes RGB, profundidad e infrarrojo. Se dividen las secuencias grabadas en una parte para entrenamiento y otra parte para evaluación. Aquellas destinadas al entrenamiento del modelo son etiquetadas manualmente para separar los materiales unos de otros.

Una vez tengamos los materiales separados, se generan descriptores de esta información mediante la transformada de *Fourier* y se crean modelos de conocimiento a partir de estos datos, usando para ello el modelo probabilístico *Gaussian Mixture Models*, de forma que tengamos un modelo por cada material.

En la parte de evaluación, se estudia la bondad de nuestros modelos usando para ello las secuencias destinadas a la parte de evaluación. Se realiza la transformada de *Fourier* de la secuencia completa y se introduce en cada uno de los modelos generados, de forma que devuelve la probabilidad que tiene cada pixel de pertenecer a un objeto.

También se estudian posibles modificaciones que se pueden realizar a los descriptores para mejorar la identificación.

Por último, se recogen las conclusiones de este trabajo y se menciona el posible trabajo futuro en relación al tema escogido.

Palabras clave

Kinect, Fourier, Frecuencia, Infrarrojo, Profundidad, Espectro visible, Video, Imagen, GMM.

Abstract

The objective of this TFG is how to use frequency information in a sequence of images to be able to automatically identify materials. In addition, it also seeks to study the impact of adding images with infrared light and depth information (other than RGB images) in the development of this task.

To do this, we start by creating a new dataset with the help of the *Kinect* sensor, composed of RGB, depth and infrared images. The recorded sequences are divided into two parts: one part for training and another part for test. Those intended for the training of the model are manually labeled to separate the materials from each other.

Once we have the materials separated, we generate descriptors of this information using the Fourier transform and create knowledge models from these data, using the probabilistic model Gaussian Mixture Models. In conclusion, we have a model for each material.

In the evaluation part, we study the goodness of our models using the sequences assigned to the evaluation part. The Fourier transform of the complete sequence is performed and introduced in each of the generated models, so that it returns the probability that each pixel has to belong to an object.

Possible modifications that can be made to descriptors to improve identification are also studied.

Finally, we collect the conclusions of this work and mention the possible future work in relation to the chosen task..

Keywords

Kinect, Fourier, Frequency, Infrared, Depth, Visible Spectrum, Video, Image, GMM.

Agradecimientos

No puedo dejar pasar la oportunidad de agradecer a aquellas personas que han estado a mi lado estos años y que me han hecho amar esta carrera como la he amado.

Para empezar, me gustaría agradecer a mi tutor Marcos, por su ayuda y su paciencia, por su confianza y su cercanía. Gracias de verdad.

Continuar con aquellos que se han convertido en mis hermanos, que han peleado a mi lado y que han aguantado bajo la lluvia como campeones: El gran Pab, Joselito, Emilin, Serrita, Sergi, Julianin... Nos espera una vida apasionante, y quiero vivirla cerca vuestro.

También necesito agradecer a mi familia, que han tenido que aguantar conmigo esta etapa de madurez, que han sabido entenderme y aconsejarme. Os quiero.

Gracias a Luci, que siempre ha estado a mi lado. Tu me conoces de verdad.

Gracias a aquellos que habéis hecho que estos años hayan sido tan bonitos como han sido: Ana, Clau, Paula... Sois lo mejor que me llevo de aquí.

Gracias Juliet, por acompañarme en este final de etapa, y confiar en mí.

Gracias señores del L.O.L. Los momentos vividos a vuestro lado son únicos.

Gracias a todos los profesores que han conseguido transmitirme su emoción y me han animado a ser mejor estudiante y persona.

Gracias a Almudena, tu me has enseñado a conocerme. Se puede decir, que he llegado con los deberes hechos.

Índice general

Resumen	v
Abstract	vii
Agradecimientos	ix
1. Introducción	1
1.1. Motivación	1
1.2. Objetivos	2
1.3. Organización de la memoria	2
2. Estado del arte	5
2.1. Introducción	5
2.2. Captura de datos	5
2.2.1. Microsoft Kinect V2	6
2.3. Dataset existentes	6
2.4. Caracterización de los materiales	8
2.4.1. Preprocesado de la información	8
2.5. Clasificadores	9
3. Creación del dataset	11
3.1. Introducción	11
3.2. Desarrollo	11
3.2.1. Captura de las imágenes	11
3.2.2. Enmascarado	14
4. Entrenamiento	17
4.1. Introducción	17
4.2. Caracterización	17
4.3. Generación de los modelos de conocimiento	18
4.4. Resumen	19
5. Evaluación	21
5.1. Introducción	21
5.2. Marco de evaluación	22
5.2.1. <i>Dataset</i>	22

5.2.2.	Conjunto de datos de entrenamiento y evaluación	22
5.2.3.	Conjunto de descriptores	23
5.3.	Resultados	23
5.3.1.	Cualitativos	23
5.3.2.	Cuantitativos	24
5.3.3.	Discusión de resultados	24
6.	Conclusiones y trabajo futuro	33
6.1.	Conclusiones	33
6.2.	Trabajo futuro	33
	Bibliografía	34
A.	Estudios publicados sobre detección de materiales	37
B.	Resto de vistas para infrarrojo y profundidad.	41

Índice de figuras

2.1. Diseño físico de la Kinect	6
2.2. Adaptador Kinect a PC	6
2.3. Ejemplos de la base de datos Flickr.	7
2.4. Ejemplos de la base de datos CURET	7
2.5. Ejemplo de materiales en la base de datos MPI-VIPS	8
3.1. Distintas posiciones de la cámara para la misma disposición	13
3.2. Vista cenital y con la luz encendida de la disposición 1	13
3.3. Vista cenital y con la luz encendida de la disposición 2	14
3.4. Ejemplos de máscaras creadas con RatSnake	15
3.5. Ejemplos de imágenes tras aplicar alguna de las máscaras.	15
4.1. Esquema para la generación del modelo de conocimiento.	17
4.2. Ejemplos de <i>AICS</i>	18
5.1. Esquema para evaluar el modelo de conocimiento.	21
5.2. Imágenes de probabilidad cenital.	29
5.3. Imágenes de probabilidad cenital. tras normalización	30
5.4. Imágenes de probabilidad cenital. tras pca	31
5.5. Media de la probabilidades de cada pixel.	32
5.6. Media de la probabilidades de cada pixel normalizados.	32
5.7. Media de la probabilidades de cada pixel tras pca.	32
A.1. Imagen explicativa del metodo del paper “Visual Viobrometry”	38
A.2. Cámara de reflectancia basada en espejo parabólico	39
A.3. Funcionamiento del aLDA completo	39
B.1. Imágenes de probabilidad frontal.	42
B.2. Imágenes de probabilidad lateral.	43
B.3. Imágenes de probabilidad pseudolateral.	44
B.4. Imágenes de probabilidad frontal tras normalización	45
B.5. Imágenes de probabilidad lateral tras normalización.	46
B.6. Imágenes de probabilidad pseudolateral tras normalización.	47
B.7. Imágenes de probabilidad frontal tras pca	48
B.8. Imágenes de probabilidad lateral tras pca.	49
B.9. Imágenes de probabilidad pseudolateral tras pca.	50

Índice de tablas

2.1. Características de las imágenes capturadas por la Kinect	6
3.1. <i>Script</i> de grabación	12
5.1. Conjunto de datos utilizados para entrenar el modelo	22
5.2. Tabla para el algodón.	25
5.3. Tabla para el cartón.	25
5.4. Tabla para la cerámica.. . . .	26
5.5. Tabla para el cristal.. . . .	26
5.6. Tabla para el metal.	27
5.7. Tabla para el plástico.	27
5.8. Tabla para el teflón.	28

Capítulo 1

Introducción

1.1. Motivación

El reconocimiento de materiales es una tarea muy interesante debido a sus múltiples y diversos posibles usos. Conocer el material de forma automática con el que está hecho un objeto, nos permitirá una toma de decisiones y una reacción en consecuencia a lo que tenemos enfrente. La fuerza que deberá aplicar un robot será distinta si lo que pretende agarrar es un ladrillo o un vaso de cristal, por ejemplo. También permitiría detectar si nos encontramos ante un objeto peligroso (como dinamita) y tomar las medidas necesarias.

Al hablar de identificar materiales y distinguir unos de otros, nos encontramos ante un reto que resulta sencillo para los seres humanos pero con una alta complejidad si queremos automatizarlo, debido entre otras cosas a que es difícil saber cuáles son las características de un material que los distinguen de otro. Los seres humanos utilizan la información del material del que está compuesto un objeto para la toma de decisiones, por lo que el diseño de un sistema artificial automático para el reconocimiento de materiales permitiría en teoría traspasar parte de estos procesos de decisión a un sistema artificial.

Conseguir una buena identificación de materiales podría suponer un gran avance en algunas industrias, como la automovilística por ejemplo, ya que la existencia de los llamados “coches autónomos” puede nutrirse de la información que le aporte distinguir materiales: saber si nos encontramos en carretera o tierra y ajustar la velocidad, reaccionar ante un obstáculo que encontremos en la carretera decidiendo si esquivarlo o no (ya que intentar evitar pasar por encima podría provocar un accidente y quizá tan solo se trate de una bolsa de plástico), etc.

Hasta el momento, la gran mayoría de tecnologías que buscan la identificación de

materiales de forma automática se basan en el uso de imágenes con información de color de cámaras estándar. Este tipo de información, depende en gran medida de la iluminación en la que nos encontremos, por lo que puede ser interesante encontrar otras fuentes de información que no varíen en presencia de cambios de iluminación. Mediante el uso del sensor *Kinect*, podremos obtener imágenes con información de profundidad y luz infrarroja (además de la de color) que son inmunes a los cambios de iluminación de la escena ya que trabajan en una frecuencia inferior a la del espectro visible.

Este trabajo intentará poner de manifiesto el gran potencial que puede tener el uso de estas imágenes con información de profundidad y luz infrarroja para la tarea requerida.

1.2. Objetivos

En este trabajo se explora el impacto del uso de información frecuencial de imágenes de color, profundidad y luz infrarroja para grabaciones de objetos sencillos, con el objetivo de poder automatizar la identificación de materiales. Para ello:

- 1) Se realiza un estudio del estado del arte para conocer los métodos existentes en la actualidad que buscan obtener una solución al problema propuesto.
- 2) Se genera de una nueva base de datos con información de color, profundidad e infrarrojo, con materiales sencillos en diferentes situaciones y bajo diferentes tipos de iluminación.
- 3) Se realiza un etiquetado de la base de datos de forma manual, distinguiendo unos materiales de otros, y división de los datos en una parte de train y otra de test.
- 4) Se realiza un preprocesado de la información obtenida de la base de datos, extracción de características frecuenciales y entrenamiento de diferentes clasificadores.
- 5) Se evalúa la bondad de estos clasificadores mediante los datos almacenados para test.

1.3. Organización de la memoria

La memoria consta de los siguientes capítulos:

- Capítulo **Introducción**.
- Capítulo **Estado del arte**.
- Capítulo **Creación del dataset**.

- Capítulo ENTRENAMIENTO.
- Capítulo Evaluación.
- Capítulo Conclusiones y trabajo futuro.
- Bibliografía.

Capítulo 2

Estado del arte

2.1. Introducción

La identificación de materiales es un tema de mucha relevancia y por consecuencia altamente investigado en la actualidad. Son muchos los distintos estudios realizados sobre el tema, por lo que es importante destacar que para la realización de este trabajo se ha tenido especial interés en aquellos que pudieran verse beneficiados del uso de información adicional de profundidad e infrarrojo, así como de aquellos que hacían alusión al uso de componentes frecuenciales como características descriptivas de los materiales.

Para nuestro trabajo, tomaremos la tarea de identificación de materiales basándonos en un sistema de reconocimiento genérico el cual consta de las siguientes etapas:

1. Se comienza con la selección del origen de los datos, normalmente desde el dataset.
2. La segunda etapa consiste en la extracción de características, en nuestro caso se hará obteniendo la transformada de Fourier del valor de los píxeles de cada material a lo largo de una secuencia de video.
3. La tercera y última etapa consiste en generar el modelo de conocimiento a partir de unos datos de entrenamiento.

2.2. Captura de datos

Para conseguir imágenes con información de color, infrarrojo y profundidad se ha hecho uso del sensor *Microsoft Kinect v2* [1] (Fig:2.1) y del adaptador de *Kinect* para



Figura 2.1: Diseño físico de la Kinect.



Figura 2.2: Adaptador para PC

PC Windows [2](Fig: 2.2). Ambos son productos fáciles de adquirir y con un precio bajo, por lo que nos resultarán óptimos para este TFG.

El adaptador nos permitirá utilizar el sensor *Kinect* desde un PC convencional, permitiéndonos tener los datos capturados directamente en la estación de trabajo.

2.2.1. Microsoft Kinect V2

Este producto, diseñado y comercializado por la empresa *Microsoft* en el año 2010, fue pensado como un controlador de videojuegos para la videoconsola *Xbox* creada por esta misma empresa. Se trata de un conjunto de cámaras dispuestas en horizontal en una barra de aproximadamente 23cm capaces de capturar imágenes en color, imágenes en infrarrojo e imágenes en profundidad con las siguientes características:

Tabla 2.1: Características de las imágenes capturadas por la Kinect

Sensor	Color	Profundidad	Infrarrojo
Resolución	1920×1080 (1080p)	512x424	512x424
FrameRate	30fps	30fps	30fps

2.3. Dataset existentes

En este apartado veremos algunos de los *dataset* de materiales existentes en la actualidad así como sus puntos fuertes y débiles para la identificación de materiales:

- Flickr Material Database[3]: imágenes a color de la superficie de diez de los materiales más comunes en la vida cotidiana (tela, follaje, cristal, cuero, me-

tal, papel, plástico, piedra, agua y madera). Existen 100 imágenes para cada categoría, 50 normales y 50 cercanas, y fueron seleccionadas manualmente para garantizar que, para cada material, aparezcan diferentes composiciones, colores, texturas etc. Figura:2.3



Figura 2.3: Ejemplos de la base de datos Flickr. Por orden de aparición de izquierda a derecha: madera, tela, follaje y cristal.

- CURET(Columbia-Utrecht Reflectance and Texture Database)[4]: En realidad se trata de tres bases de datos diferenciadas. En la primera tenemos una base de datos BRDF (función de distribución de reflectancia bidireccional) con mediciones de reflectancia para más de 60 muestras diferentes, cada una observada con más de 200 combinaciones diferentes de direcciones de visualización e iluminación. La segunda consiste en una base de datos de BRDF con parámetros de ajuste que pueden ser utilizados directamente tanto para el análisis como para la síntesis de imágenes. Por último, la tercera base de datos esta formada por datos BTF (bidirectional texture function) con texturas de imágenes de más de 60 muestras diferentes, cada una observada con más de 200 combinaciones diferentes de direcciones de visualización e iluminación.Figura:2.4. Es interesante observar que en estas bases de datos están basadas en la reflectancia que muestran los objetos, muy en concordancia con los objetivos de este TFG.

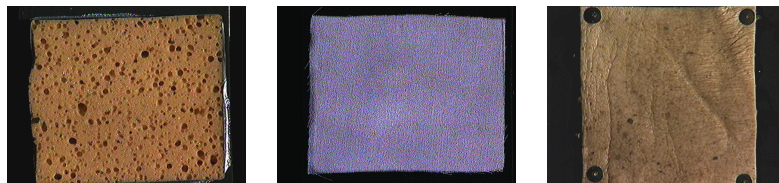


Figura 2.4: Ejemplos de la base de datos CURET. De izquierda a derecha: esponja, poliéster, piel humana.

- MPI-VIPS[5]: Se trata de un dataset bastante completo, con 11 diferentes clases de materiales, cada uno de ellos observado desde diferentes puntos de vista y con cambios de iluminación. Figura: 2.5



Figura 2.5: Ejemplo de materiales en la base de datos MPI-VIPS. De izquierda a derecha: aluminio, hoja de una planta, madera.

2.4. Caracterización de los materiales

Para la caracterización de los materiales usaremos la transformada de *Fourier* (Ecuación: 2.1), que nos permitirá extraer los componentes frecuenciales de cada material (Más sobre el uso de frecuencia para caracterizar materiales en el apéndice: A). Más concretamente, realizaremos esta transformada sobre cada uno de los píxeles de los diferentes materiales a lo largo de una secuencia de 37 imágenes (para ello, en la fase de entrenamiento, se tendrá que realizar un enmascarado de cada una de las imágenes individualmente y para cada material que se explicará en la sección: 3.2.2).

$$g(\xi) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(x) e^{-i\xi x} dx \quad (2.1)$$

$$X_k = \sum_{n=0}^{N-1} x_n e^{-i2\pi k \frac{n}{N}} \quad k = 0, \dots, N-1. \quad (2.2)$$

2.4.1. Preprocesado de la información

Una vez hallamos obtenido y tengamos almacenados los descriptores de los materiales, será necesario realizar un pre-procesamiento de los datos debido a su gran dimensionalidad. Para ello, utilizaremos un Análisis de componentes frecuenciales o PCA [6]. Consiste en una técnica que, a través de una transformación lineal, busca cambiar el sistema de coordenadas en el que nos encontramos a uno donde los datos proyectados queden mejor representados. Este método supone que la información más relevante de los datos se encuentra en los componentes con alta varianza por lo

que se pueden ignorar aquellos con varianza baja. Los nuevos datos obtenidos estarán decorrelados entre si.[7]

2.5. Clasificadores

A la hora de generar el modelo de conocimiento, la elección de los clasificadores que vayan a ser utilizados sobre nuestros datos es una tarea crítica. Cada uno de los clasificadores existentes tiene ventajas y desventajas sobre el resto, por lo que es muy complicado saber cual será la elección óptima. Por ello, en este trabajo se ha escogido estudiar varios de estos clasificadores para después evaluar cual sería el clasificador que mejor se adapta a nuestro problema:

- *Random Forest*[8]: Se basan en el uso de múltiples árboles de decisión([9],[10]), contruidos cada uno con varias de las variables aleatorias de nuestro conjunto de entrenamiento, para realizar una clasificación. Cuando llega un dato de entrada, cada uno de los árboles creados devolverá la clase a la que predice que pertenece el dato y la clase final se calculará mediante un sistema de votación.
- *Support Vector Machines*[11]: Se trata de conseguir un hiperplano o conjunto de hiperplanos que dividan a los datos en diferentes clases minimizando una función que marca la distancia de los ejemplos más cercanos de cada clase a ese hiperplano.
- *K-means*[12]: Clasificador que funciona en cuatro etapas:
 1. Inicialización: Se seleccionan N vectores como los centroides iniciales.
 2. Búsqueda de los vecinos más cercanos: Para cada dato de entrenamiento, buscaremos el centroide más cercano y asociaremos dicho dato a la clase correspondiente representada por el centroide.
 3. Actualizar centroides: Calculamos nuevos centroides con los datos asociados de la etapa anterior.
 4. Repetir 2. y 3. hasta que la distancia media sea menor que un umbral establecido.
- *Nearest neighbor*[13]: Dado un nuevo dato de entrada, se buscará la clase a la que pertenece dependiendo de cual sea el dato más cercano a el.
- *Gaussian Mixture Models*[14]: Utilizaremos este modelo probabilístico basado en el cálculo de medias y covarianzas con nuestros datos de entrenamiento para encontrar las funciones gaussianas que mejor se adapten a ellos.

Capítulo 3

Creación del dataset

3.1. Introducción

Como se ha podido observar en el apartado anterior, en la actualidad no existe ninguna base de datos de materiales que incorpore entre sus datos imágenes de infrarrojo y profundidad de cada uno de estos materiales. Por ello, uno de los objetivos de este TFG ha sido el de crear una base de datos que incorpore esta información. Además, será con esa misma base de datos con la que posteriormente se ha trabajado para entrenar y evaluar el modelo de conocimiento generado.

Por lo tanto, en este apartado veremos en profundidad los pasos que se han seguido para crear el dataset.

3.2. Desarrollo

3.2.1. Captura de las imágenes

Como ya se ha introducido en apartados anteriores, para la captura de imágenes se ha hecho uso de una cámara *Microsoft Kinect V2* y de un adaptador desde la cámara a un PC (Figura:2.2). Además, se ha hecho uso de un trípode en el que poder situar la cámara en diferentes lugares dependiendo de lo que quisiéramos capturar. La grabación se ha realizado en el laboratorio *Video Processing and Understanding Lab (VPU)* de la Escuela Politécnica Superior de la Universidad Autónoma de Madrid.

Para intentar que la base de datos sea lo más completa posible, se han seleccionado varias disposiciones de los materiales, varias posiciones de la cámara y diferentes iluminaciones. Con diferentes combinaciones de estas situaciones podremos obtener un gran cantidad de imágenes para nuestro *dataset*.

Se ha escogido realizar 16 grabaciones de secuencias diferentes, las cuales dividi-

remos en 8 como muestras de entrenamiento y 8 como muestras de evaluación. Para las imágenes de entrenamiento tendremos una disposición de los materiales y para las de test tendremos otra. Las posibles posiciones de la cámara serán: vista cenital, vista frontal, vista lateral y vista pseudolateral (Ejemplo de estos puntos de vista para la disposición 1 en la figura: 3.1).

Además para cada una de estas vistas, grabaremos una secuencia en la que las luces estarán apagadas y otra en la que estén encendidas. En resumen, las 16 secuencias quedan de la siguiente forma: tabla 3.1, donde la disposición 1 estará formada por los siguientes materiales: cartón, cristal (copa), algodón, vidrio, plástico (soporte del trípode, cubo de *rubik* y carpeta), cerámica (plato), teflón (sartén) y metal (podemos ver un ejemplo de esta disposición desde una vista cenital y con la luz encendida en la figura:3.2) y la disposición 2 estará formada por los mismos objetos, colocados en la misma posición, pero añadiremos otros objetos nuevos hechos con materiales ya presentes en la primera disposición: incluimos otra sartén de teflón, otra carpeta de plástico, mas objetos de metal, un nuevo fragmento de cartón y algunas muestras más de plástico (un ejemplo de esta disposición en vista cenital y con la luz encendida lo podemos encontrar en la figura: 3.3)

Escena	Disposición	Posición	Luz
1	1	Frontal	Encendida
2	1	Frontal	Apagada
3	1	Lateral	Encendida
4	1	Lateral	Apagada
5	1	Cenital	Encendida
6	1	Cenital	Apagada
7	1	Pseudolateral	Encendida
8	1	Pseudolateral	Apagada
9	2	Frontal	Encendida
10	2	Frontal	Apagada
11	2	Lateral	Encendida
12	2	Lateral	Apagada
13	2	Cenital	Encendida
14	2	Cenital	Apagada
15	2	Pseudolateral	Encendida
16	2	Pseudolateral	Apagada

Tabla 3.1: *Script* de grabación



Figura 3.1: Observamos la misma disposición para los cuatro diferentes puntos de vista posibles en color. De izquierda a derecha y de arriba a abajo: cenital, frontal, lateral y pseudo-lateral.

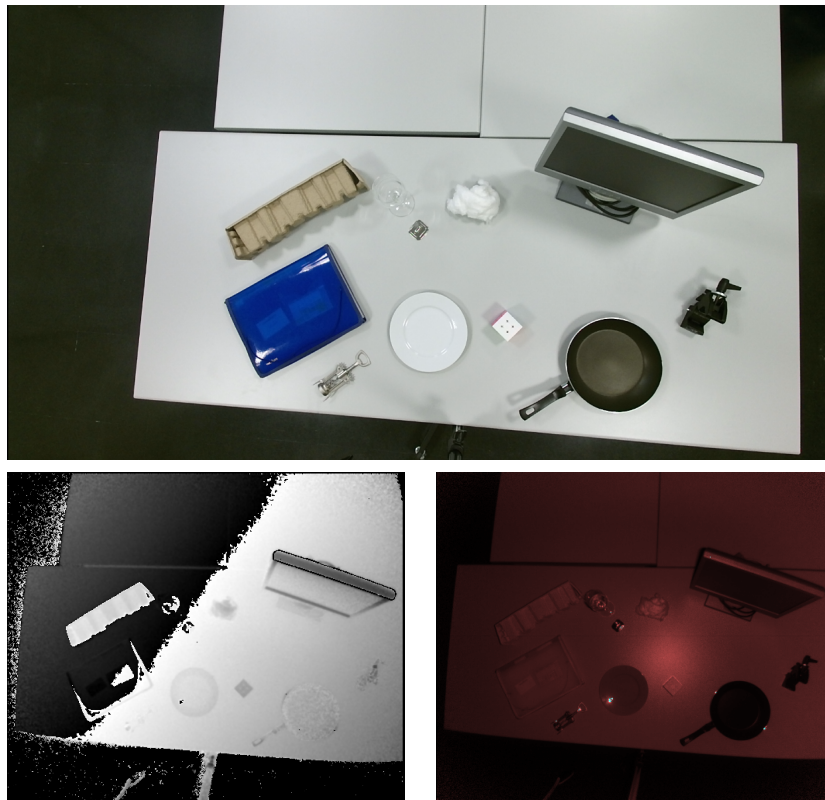


Figura 3.2: Vista cenital y con la luz encendida de la disposición 1 en color, profundidad e infrarrojo.

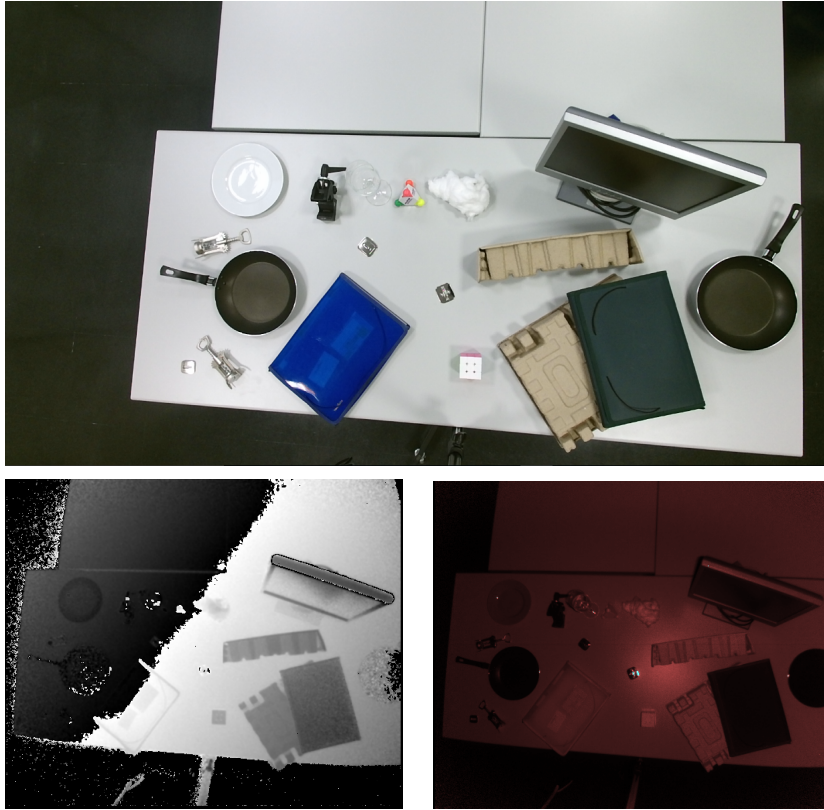


Figura 3.3: Vista cenital y con la luz encendida de la disposición 2 en color, profundidad e infrarrojo.

3.2.2. Enmascarado

Una vez hemos grabado las secuencias de nuestra base de datos, debemos proceder a efectuar un enmascarado manual de todos los objetos presentes en cada una de las escenas. Para ello se ha hecho uso del programa *RatSnake* [15], una herramienta de anotación manual que nos permite realizar la tarea que buscamos de una forma sencilla y relativamente rápida. Algunos ejemplos de máscaras creadas los podemos ver en la figura: 3.4.

Al aplicar estas máscaras sobre nuestros vídeos, podremos aislar todos los píxeles de un objeto a lo largo de cada una de las secuencias. Gracias a ello, nuestra base de datos quedará dividida en distintos materiales. Podemos ver un ejemplo de una imagen tras aplicar algunas de estas máscaras en 3.5

Para cada uno de estos materiales tendremos almacenado como varían todos los píxeles de todas las secuencias en los que aparece dicho material a lo largo de cada vídeo.

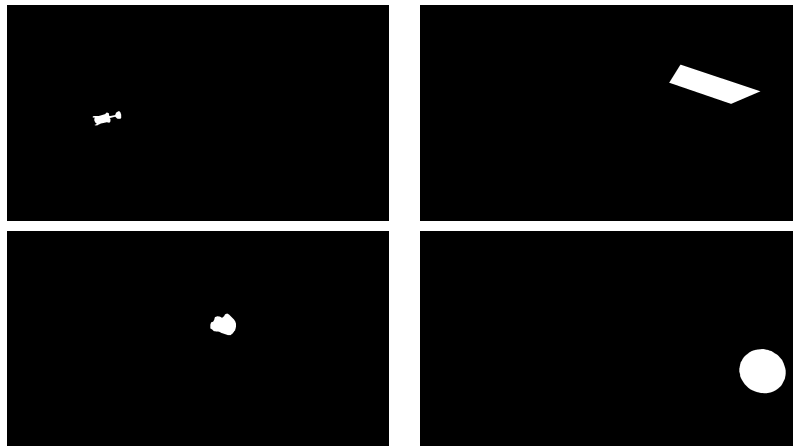


Figura 3.4: Ejemplos de máscaras creadas con RatSnake (en este caso desde vista cenital para la disposición 2, pero se ha realizado para todas las secuencias). De izquierda a derecha y de arriba a abajo: sacacorchos, pantalla de ordenador, algodón y teflón.

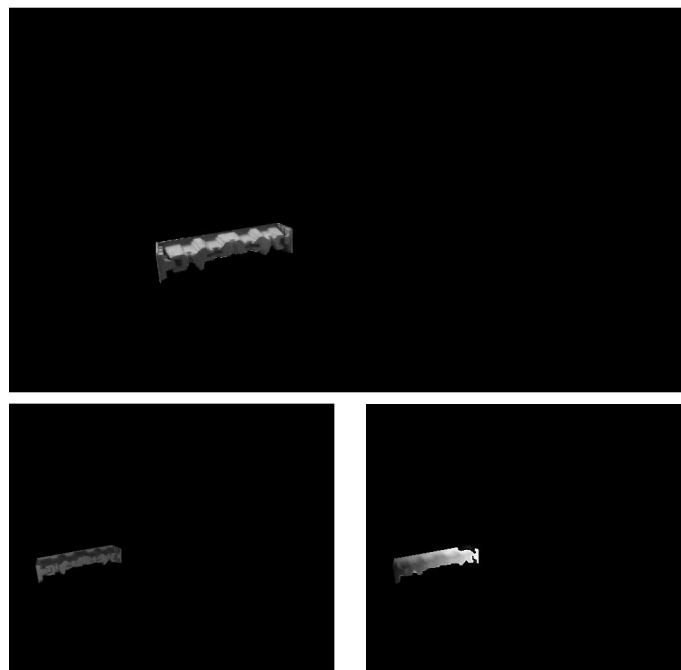


Figura 3.5: Ejemplos de imágenes tras aplicar alguna de las máscaras creadas. Se observa un fragmento de cartón desde el punto de vista frontal y separado en información de color, infrarrojo y profundidad.

Capítulo 4

Entrenamiento

4.1. Introducción

En este apartado se explica en detalle el esquema que se ha seguido en este trabajo para la generación del modelo de conocimiento (figura 4.1)

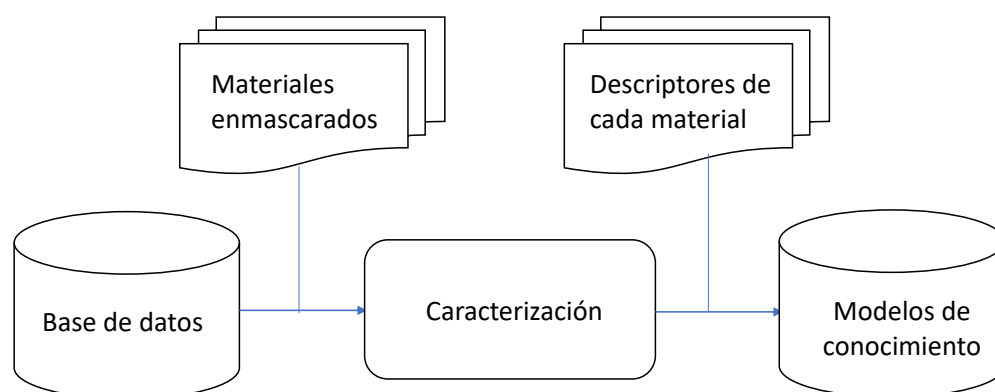


Figura 4.1: Esquema explicativo de la generación del modelo de conocimiento.

4.2. Caracterización

En la base de datos la información está distribuida de forma que, para cada material, están almacenados todos los píxeles en los que aparece a lo largo de las secuencias grabadas. Una vez extraída esta información, será necesario obtener descriptores de

estos datos que nos permitan caracterizar la forma en la que se comportan, buscando alguna característica que píxeles de un mismo material puedan compartir y que no compartan píxeles de materiales distintos.

Con este objetivo, en este trabajo se ha escogido obtener los descriptores de cada material como la transformada de *Fourier* 2.1 de cada uno de los píxeles a lo largo de la secuencia en la que ha sido grabada, de forma que utilicemos como caracterización los coeficientes frecuenciales de variación del valor del pixel. Estos serán los descriptores con los que se generan los modelos de conocimiento.

4.3. Generación de los modelos de conocimiento

Usaremos el modelo probabilístico “*Gaussian Mixture Model*” que nos permitirá construir un clasificador para cada material. Este modelo buscará ajustar un número de gaussianas (el cual es elegido por el diseñador) que mejor se adapte a nuestros datos de entrada. De esta forma, si por ejemplo decidimos calcular 5 gaussianas, obtendremos 5 medias diferentes y 5 matrices de covarianza.

Es decir, cada conjunto de transformadas del valor del píxel de cada material tendrá asociado un clasificador GMM del orden que seleccionemos. Para conocer cual será el número de gaussianas óptimo se ha hecho uso del índice *AIC* (*Akaike: akaike information criterion*) el cual nos indica como de óptima es la separación entre estas gaussianas: cuanto menor sea, mejor adaptados a los datos de entrenamiento estará nuestro clasificador. Para nuestro caso en concreto, los descriptores de los materiales respondían mejor seleccionando dos gaussianas (en la figura 4.2 se puede observar algunos ejemplos).

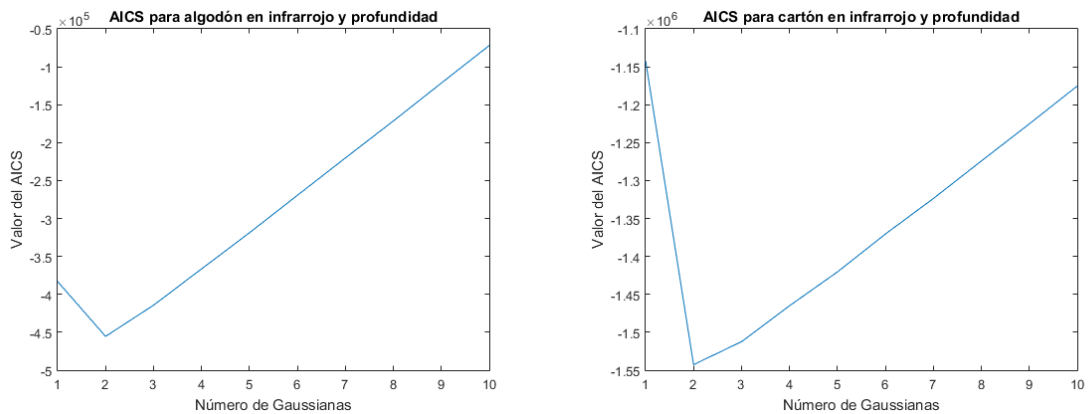


Figura 4.2: Ejemplos de *AICS*. Se observa como el menor *AICS* aparece cuando seleccionamos 2 Gaussianas para modelar nuestros datos.

4.4. Resumen

A modo de resumen, los pasos que se han seguido para generar los modelos son los siguientes:

1. Extraemos los datos de la base de datos.
2. Para cada material (por ejemplo algodón), calculamos la transformada de *Fourier* de todos sus píxeles y lo almacenamos en una matriz (de forma que tengamos, siguiendo con el ejemplo, la transformada de todos los píxeles pertenecientes al algodón de todas las secuencias). Si hacemos esto mismo para el resto de materiales, tendremos una matriz por material. Estos son los descriptores con los que entrenamos el modelo.
3. Calculamos el GMM de cada material utilizando los descriptores de forma que tengamos un modelo para cada material.

Capítulo 5

Evaluación

5.1. Introducción

En este apartado se profundizará en las diferentes técnicas que se han realizado para evaluar la bondad de nuestro modelo de conocimiento y se discutirán los resultados obtenidos. El esquema que se ha seguido en esta parte del trabajo es el siguiente (Figura: 5.1).

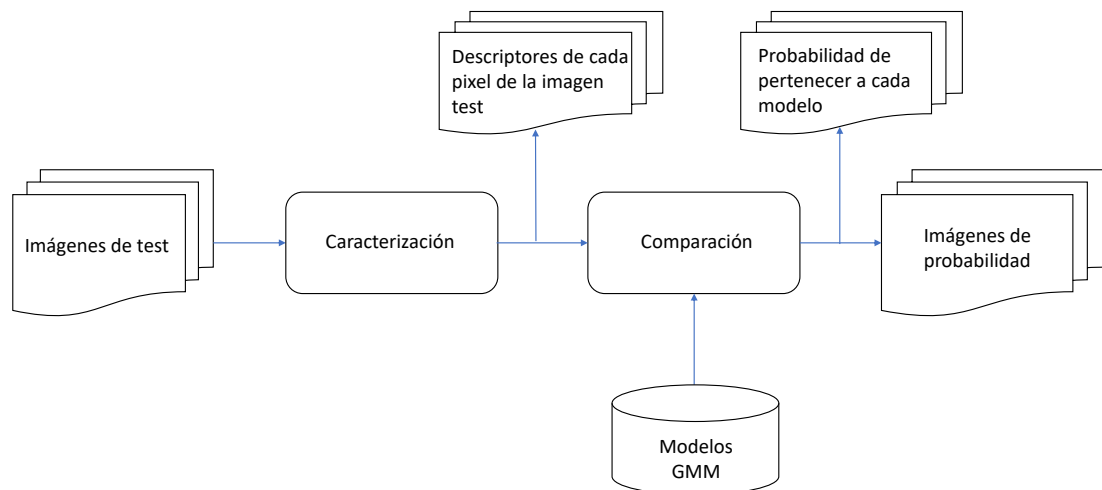


Figura 5.1: Esquema que se ha seguido para evaluar el modelo de conocimiento ya entrenado.

5.2. Marco de evaluación

5.2.1. Dataset

Para evaluar nuestro modelo hemos utilizado como datos de evaluación la mitad de las secuencias grabadas con la *Kinect* (ver: 3), las cuales son muy similares a las utilizadas en entrenamiento pero añadiendo más objetos en diferentes posiciones.

Además, se ha realizado una construcción de máscaras sobre las imágenes de evaluación a modo de *ground truth*.

5.2.2. Conjunto de datos de entrenamiento y evaluación

Debido a la gran dimensionalidad de la información de las imágenes en color (principalmente por su alta resolución: 1920×1080) añadido a que el objetivo de este trabajo es averiguar como de útil puede ser el uso de imágenes con información de infrarrojo y profundidad para el reconocimiento de materiales, se ha decidido no tener en cuenta la información RGB en este capítulo y se tendrá en cuenta como trabajo futuro.

De esta forma tendremos el siguiente conjunto de datos:

- Entrenamiento¹:

Material	Número de muestras
Algodón	6560 píxeles
Cartón	19146 píxeles
Cerámica	16648 píxeles
Cristal	7358 píxeles
Metal	6828 píxeles
Plástico	49718 píxeles
Teflón	31530 píxeles
Extra	156126 píxeles

Tabla 5.1: Conjunto de datos utilizados para entrenar el modelo

- Evaluación: Disponemos de 8 secuencias de 37 imágenes cada una para evaluar nuestro modelo, todas grabando la misma situación pero con cambios de punto de vista e iluminación. Las imágenes de profundidad e infrarrojo tienen resolución 512×424 , lo que implica que al realizar la caracterización de toda la secuencia tendremos un total de 217088 descriptores por secuencia.

¹La clase extra ha sido calculada con un *sampling* de todos aquellos píxeles que no pertenecen a una de las clases anteriores.

5.2.3. Conjunto de descriptores

Además de los descriptores iniciales, los cuales consisten en una transformada de *Fourier* sobre los datos en crudo, se han implementado procesos previos y posteriores al cálculo de estas características, con el fin de intentar mejorar los resultados obtenidos. Los dos nuevos cálculos implementados son los siguientes:

- Normalización basada en la varianza: Se realiza una normalización previa de los descriptores mediante la media (\bar{x}) y la desviación estándar (σ), buscando que los nuevos descriptores tengan $\bar{x} = 0$ y $\sigma = 1$ siguiendo la ecuación 5.1, y será con estos descriptores normalizados con el que se generarán los modelos de los materiales.

$$descriptores_{normalizados} = \frac{descriptores - \bar{x}}{\sigma} \quad (5.1)$$

- Preprocesado PCA: En este caso, se realiza un preprocesado de la información mediante PCA (**Preprocesado de la información**) de forma que construyamos el modelo teniendo en cuenta solamente los principales componentes de nuestros descriptores, reduciendo complejidad y, presumiblemente, mejorando los resultados. Además, a estos nuevos descriptores también se le podrá realizar la normalización previa que se ha mencionado.

5.3. Resultados

5.3.1. Cualitativos

Se comienza con la extracción de descriptores de las secuencias de test completas. Estos descriptores se introducen en cada uno de los 9 modelos creados (8 materiales y 1 clase extra), de forma que se obtenga a la salida la probabilidad de cada pixel de pertenecer a cada modelo. Debido a que estamos trabajando en la banda de infrarrojos y con información de profundidad, los cambios en la iluminación no deberían surtir efecto sobre nuestra evaluación, por lo que obtendremos los mismos resultados con luz apagada y luz encendida.

A continuación se representan imágenes de probabilidad desde vista cenital y luz apagada, que son el resultado de haber pasado los descriptores de cada secuencia de test por nuestros modelos (ver Figura: 5.2 | Resto de figuras en??)

Otra forma de observar estos datos de forma cualitativa es mediante la asignación de etiquetas por minimización de energías. En la Figura 5.5 podemos ver un ejemplo desde la vista cenital de como quedarían estas nuevas imágenes.

- Primera mejora: Intentaremos ahora mejorar estos resultados mediante la normalización mencionada en 5.2.3. Nuestras nuevas imágenes de probabilidad son las siguientes 5.3

De nuevo, podremos representarlo mediante la minimización de energías de cada pixel 5.6

- Segunda mejora: Aplicaremos un preprocesado *PCA* 5.2.3 para intentar mejorar los resultados, teniendo en cuenta solo los componentes principales de nuestros descriptores.

Tras la ejecución de este preprocesado obtenemos las nuevas imágenes de probabilidad 5.4

Y al igual que en los casos anteriores, también podemos calcular minimización de energías para obtener una imagen diferenciada en clases. Por ejemplo, para la vista cenital tenemos Figura: 5.7

A nivel cualitativo, las detecciones parecen empeorar cuando realizamos el normalizado y sin embargo parecen tener una alta mejoría al aplicar *PCA*. Esto podremos comprobarlo con el estudio de los resultados a nivel cuantitativo.

5.3.2. Cuantitativos

Para valorar los resultados a nivel cualitativo crearemos una matriz de resultados *precisión*, *recall* y *F-measure* a partir de los parámetros *true positives*, *false positives* y *false negatives* para cada material.

5.3.3. Discusión de resultados

A la vista de los resultados, no hemos sido capaces de distinguir cual de las tres maneras de calcular nuestros descriptores (estándar, con normalización o con preprocesado previo) es la más óptima para la detección de materiales, ya que ninguno da resultados lo suficientemente mejores que el resto para todas, o la mayoría, de las situaciones.

Además, los números en las matrices calculadas son muy bajos lo que implica, que el modelo necesitaría de mejoras para funcionar de una manera aceptable.

Aun así, algunos resultados obtenidos son esperanzadores y pueden indicar que el análisis frecuencial de secuencias y el uso de imágenes de profundidad e infrarrojo es un camino interesante para conseguir una buena identificación de materiales.

Vista	Inf/Prof	Luz	Estándar			Normalizado			PCA		
			P	R	F	P	R	F	P	R	F
Cenital	Inf	Apagada	0.03	0.89	0.06	0.01	0.98	0.03	0.02	0.98	0.04
Cenital	Inf	Encendida	0.02	0.88	0.05	0.01	0.95	0.03	0.01	0.83	0.03
Cenital	Prof	Apagada	0.03	0.89	0.06	0.01	0.97	0.03	0.02	0.97	0.04
Cenital	Prof	Encendida	0.02	0.89	0.05	0.01	0.95	0.03	0.01	0.82	0.03
Frontal	Inf	Apagada	0.01	0.53	0.03	0.00	0.05	0.00	0.00	0.32	0.01
Frontal	Inf	Encendida	0.01	0.53	0.03	0.01	0.92	0.02	0.01	0.54	0.02
Frontal	Prof	Apagada	0.01	0.53	0.03	0.00	0.06	0.00	0.00	0.28	0.01
Frontal	Prof	Encendida	0.01	0.53	0.03	0.01	0.93	0.02	0.01	0.56	0.02
Lateral	Inf	Apagada	0.00	0.04	0.00	0.00	0.27	0.01	0.00	0.01	0.00
Lateral	Inf	Encendida	0.00	0.04	0.00	0.01	0.64	0.02	0.00	0.01	0.00
Lateral	Prof	Apagada	0.00	0.03	0.00	0.00	0.26	0.00	0.00	0.00	0.00
Lateral	Prof	Encendida	0.00	0.03	0.00	0.01	0.65	0.02	0.00	0.01	0.00
Pseudolateral	Inf	Apagada	0.00	0.02	0.00	0.01	0.17	0.01	0.00	0.00	0.00
Pseudolateral	Inf	Encendida	0.00	0.02	0.00	0.00	0.43	0.00	0.00	0.00	0.00
Pseudolateral	Prof	Apagada	0.50	0.02	0.04	0.50	0.09	0.15	0.00	0.00	0.00
Pseudolateral	Prof	Encendida	0.50	0.02	0.04	0.50	0.38	0.43	0.00	0.00	0.00

Tabla 5.2: Tabla para el algodón.

Vista	Inf/Prof	Luz	Estándar			Normalizado			PCA		
			P	R	F	P	R	F	P	R	F
Cenital	Inf	Apagada	0.04	0.45	0.08	0	0	0	0.18	0.93	0.30
Cenital	Inf	Encendida	0.04	0.45	0.08	0	0	0	0.00	0.00	0.00
Cenital	Prof	Apagada	0.04	0.45	0.08	0	0	0	0.18	0.92	0.30
Cenital	Prof	Encendida	0.04	0.45	0.08	0	0	0	0.00	0.00	0.00
Frontal	Inf	Apagada	0.03	0.04	0.03	0	0	0	0.26	0.09	0.13
Frontal	Inf	Encendida	0.03	0.04	0.03	0	0	0	0.10	0.17	0.13
Frontal	Prof	Apagada	0.03	0.04	0.03	0	0	0	0.23	0.08	0.12
Frontal	Prof	Encendida	0.03	0.04	0.03	0	0	0	0.12	0.20	0.15
Lateral	Inf	Apagada	0.12	0.39	0.18	0	0	0	0.00	0.00	0.00
Lateral	Inf	Encendida	0.12	0.40	0.18	0	0	0	0.12	0.49	0.20
Lateral	Prof	Apagada	0.12	0.39	0.18	0	0	0	0.00	0.00	0.00
Lateral	Prof	Encendida	0.12	0.40	0.18	0	0	0	0.12	0.50	0.20
Pseudolateral	Inf	Apagada	0.22	0.34	0.27	0	0	0	0.00	0.00	0.00
Pseudolateral	Inf	Encendida	0.23	0.35	0.27	0	0	0	0.00	0.00	0.00
Pseudolateral	Prof	Apagada	0.50	0.34	0.40	0	0	0	0.00	0.00	0.00
Pseudolateral	Prof	Encendida	0.50	0.34	0.41	0	0	0	0.00	0.00	0.00

Tabla 5.3: Tabla para el cartón.

Vista	Inf/Prof	Luz	Estándar			Normalizado			PCA		
			P	R	F	P	R	F	P	R	F
Cenital	Inf	Apagada	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Cenital	Inf	Encendida	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Cenital	Prof	Apagada	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Cenital	Prof	Encendida	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Frontal	Inf	Apagada	0.02	0.84	0.05	0.00	0.00	0.00	0.00	0.00	0.00
Frontal	Inf	Encendida	0.02	0.84	0.05	0.00	0.00	0.00	0.00	0.00	0.00
Frontal	Prof	Apagada	0.02	0.85	0.05	0.00	0.00	0.00	0.00	0.00	0.00
Frontal	Prof	Encendida	0.02	0.85	0.05	0.00	0.00	0.00	0.00	0.00	0.00
Pseudolateral	Inf	Apagada	0.01	0.05	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Pseudolateral	Inf	Encendida	0.01	0.05	0.01	0.00	0.00	0.00	0.11	0.31	0.16
Pseudolateral	Prof	Apagada	0.50	0.05	0.10	0.00	0.00	0.00	0.00	0.00	0.00
Pseudolateral	Prof	Encendida	0.50	0.05	0.09	0.00	0.00	0.00	0.50	0.31	0.38

Tabla 5.4: Tabla para la cerámica..

Vista	Inf/Prof	Luz	Estándar			Normalizado			PCA		
			P	R	F	P	R	F	P	R	F
Cenital	Inf	Apagada	0.00	0.02	0.00	0.01	0.91	0.01	0.01	0.90	0.02
Cenital	Inf	Encendida	0.00	0.02	0.00	0.01	0.89	0.01	0.01	0.91	0.01
Cenital	Prof	Apagada	0.00	0.01	0.00	0.01	0.92	0.01	0.01	0.91	0.02
Cenital	Prof	Encendida	0.00	0.01	0.00	0.01	0.89	0.01	0.01	0.92	0.01
Frontal	Inf	Apagada	0.01	0.33	0.02	0.01	0.98	0.01	0.01	0.33	0.02
Frontal	Inf	Encendida	0.01	0.25	0.01	0.01	0.98	0.02	0.00	0.00	0.00
Frontal	Prof	Apagada	0.01	0.34	0.02	0.01	0.97	0.01	0.01	0.30	0.02
Frontal	Prof	Encendida	0.01	0.26	0.01	0.01	0.98	0.02	0.00	0.00	0.00
Lateral	Inf	Apagada	0.00	0.40	0.00	0.00	1.00	0.00	0.00	0.33	0.00
Lateral	Inf	Encendida	0.00	0.25	0.00	0.00	0.99	0.00	0.00	0.00	0.00
Lateral	Prof	Apagada	0.00	0.44	0.00	0.00	1.00	0.00	0.00	0.37	0.00
Lateral	Prof	Encendida	0.00	0.26	0.00	0.00	1.00	0.00	0.00	0.00	0.00
Pseudolateral	Inf	Apagada	0.00	0.59	0.01	0.00	1.00	0.01	0.00	0.55	0.01
Pseudolateral	Inf	Encendida	0.00	0.71	0.01	0.00	0.99	0.01	0.00	0.00	0.00
Pseudolateral	Prof	Apagada	0.50	0.58	0.54	0.50	1.00	0.67	0.50	0.53	0.51
Pseudolateral	Prof	Encendida	0.50	0.71	0.59	0.50	0.98	0.66	0.00	0.00	0.00

Tabla 5.5: Tabla para el cristal..

Vista	Inf/Prof	Luz	Estándar			Normalizado			PCA		
			P	R	F	P	R	F	P	R	F
Cenital	Inf	Apagada	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Cenital	Inf	Encendida	0.00	0.00	0.00	0.08	0.01	0.02	0.00	0.00	0.00
Cenital	Prof	Apagada	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Cenital	Prof	Encendida	0.00	0.00	0.00	0.09	0.01	0.02	0.00	0.00	0.00
Frontal	Inf	Apagada	0.00	0.00	0.00	0.01	0.03	0.01	0.00	0.00	0.00
Frontal	Inf	Encendida	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
Frontal	Prof	Apagada	0.00	0.00	0.00	0.01	0.03	0.01	0.00	0.00	0.00
Frontal	Prof	Encendida	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
Lateral	Inf	Apagada	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Lateral	Inf	Encendida	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Lateral	Prof	Apagada	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Lateral	Prof	Encendida	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Pseudolateral	Inf	Apagada	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
Pseudolateral	Inf	Encendida	0.00	0.00	0.00	0.01	0.17	0.01	0.00	0.00	0.00
Pseudolateral	Prof	Apagada	0.00	0.00	0.00	0.50	0.01	0.03	0.00	0.00	0.00
Pseudolateral	Prof	Encendida	0.00	0.00	0.00	0.50	0.17	0.25	0.00	0.00	0.00

Tabla 5.6: Tabla para el metal.

Vista	Inf/Prof	Luz	Estándar			Normalizado			PCA		
			P	R	F	P	R	F	P	R	F
Cenital	Inf	Apagada	0.04	0.01	0.02	0.00	0.00	0.00	0.00	0.00	0.00
Cenital	Inf	Encendida	0.04	0.01	0.02	0.00	0.00	0.00	0.00	0.00	0.00
Cenital	Prof	Apagada	0.04	0.01	0.02	0.00	0.00	0.00	0.00	0.00	0.00
Cenital	Prof	Encendida	0.04	0.01	0.02	0.00	0.00	0.00	0.00	0.00	0.00
Frontal	Inf	Apagada	0.18	0.13	0.15	0.00	0.00	0.00	0.11	0.30	0.16
Frontal	Inf	Encendida	0.17	0.13	0.15	0.00	0.00	0.00	0.00	0.00	0.00
Frontal	Prof	Apagada	0.18	0.13	0.15	0.00	0.00	0.00	0.11	0.32	0.16
Frontal	Prof	Encendida	0.17	0.13	0.15	0.00	0.00	0.00	0.00	0.00	0.00
Lateral	Inf	Apagada	0.04	0.10	0.06	0.00	0.00	0.00	0.00	0.00	0.00
Lateral	Inf	Encendida	0.04	0.10	0.06	0.00	0.00	0.00	0.00	0.00	0.00
Lateral	Prof	Apagada	0.04	0.09	0.05	0.00	0.00	0.00	0.00	0.00	0.00
Lateral	Prof	Encendida	0.04	0.09	0.05	0.00	0.00	0.00	0.00	0.00	0.00
Pseudolateral	Inf	Apagada	0.04	0.07	0.05	0.00	0.00	0.00	0.18	0.17	0.18
Pseudolateral	Inf	Encendida	0.04	0.07	0.05	0.00	0.00	0.00	0.00	0.00	0.00
Pseudolateral	Prof	Apagada	0.50	0.08	0.13	0.00	0.00	0.00	0.50	0.17	0.26
Pseudolateral	Prof	Encendida	0.50	0.08	0.13	0.50	0.00	0.00	0.00	0.00	0.00

Tabla 5.7: Tabla para el plástico.

Vista	Inf/Prof	Luz	Estándar			Normalizado			PCA		
			P	R	F	P	R	F	P	R	F
Cenital	Inf	Apagada	0.73	0.22	0.34	0.00	0.00	0.00	0.01	0.01	0.01
Cenital	Inf	Encendida	0.73	0.22	0.34	0.00	0.00	0.00	0.00	0.00	0.00
Cenital	Prof	Apagada	0.72	0.22	0.34	0.00	0.00	0.00	0.01	0.01	0.01
Cenital	Prof	Encendida	0.72	0.22	0.34	0.00	0.00	0.00	0.00	0.00	0.00
Frontal	Inf	Apagada	0.06	0.14	0.09	0.00	0.00	0.00	0.00	0.01	0.00
Frontal	Inf	Encendida	0.07	0.14	0.09	0.00	0.00	0.00	0.03	0.92	0.07
Frontal	Prof	Apagada	0.07	0.15	0.09	0.00	0.00	0.00	0.01	0.05	0.01
Frontal	Prof	Encendida	0.07	0.15	0.10	0.00	0.00	0.00	0.03	0.92	0.07
Lateral	Inf	Apagada	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.45	0.06
Lateral	Inf	Encendida	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.94	0.11
Lateral	Prof	Apagada	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.45	0.06
Lateral	Prof	Encendida	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.94	0.11
Pseudolateral	Inf	Apagada	0.08	0.23	0.12	0.00	0.00	0.00	0.04	0.76	0.07
Pseudolateral	Inf	Encendida	0.09	0.23	0.13	0.00	0.00	0.00	0.04	0.98	0.08
Pseudolateral	Prof	Apagada	0.50	0.22	0.31	0.00	0.00	0.00	0.50	0.80	0.61
Pseudolateral	Prof	Encendida	0.50	0.22	0.31	0.00	0.00	0.00	0.50	0.98	0.66

Tabla 5.8: Tabla para el teflón.

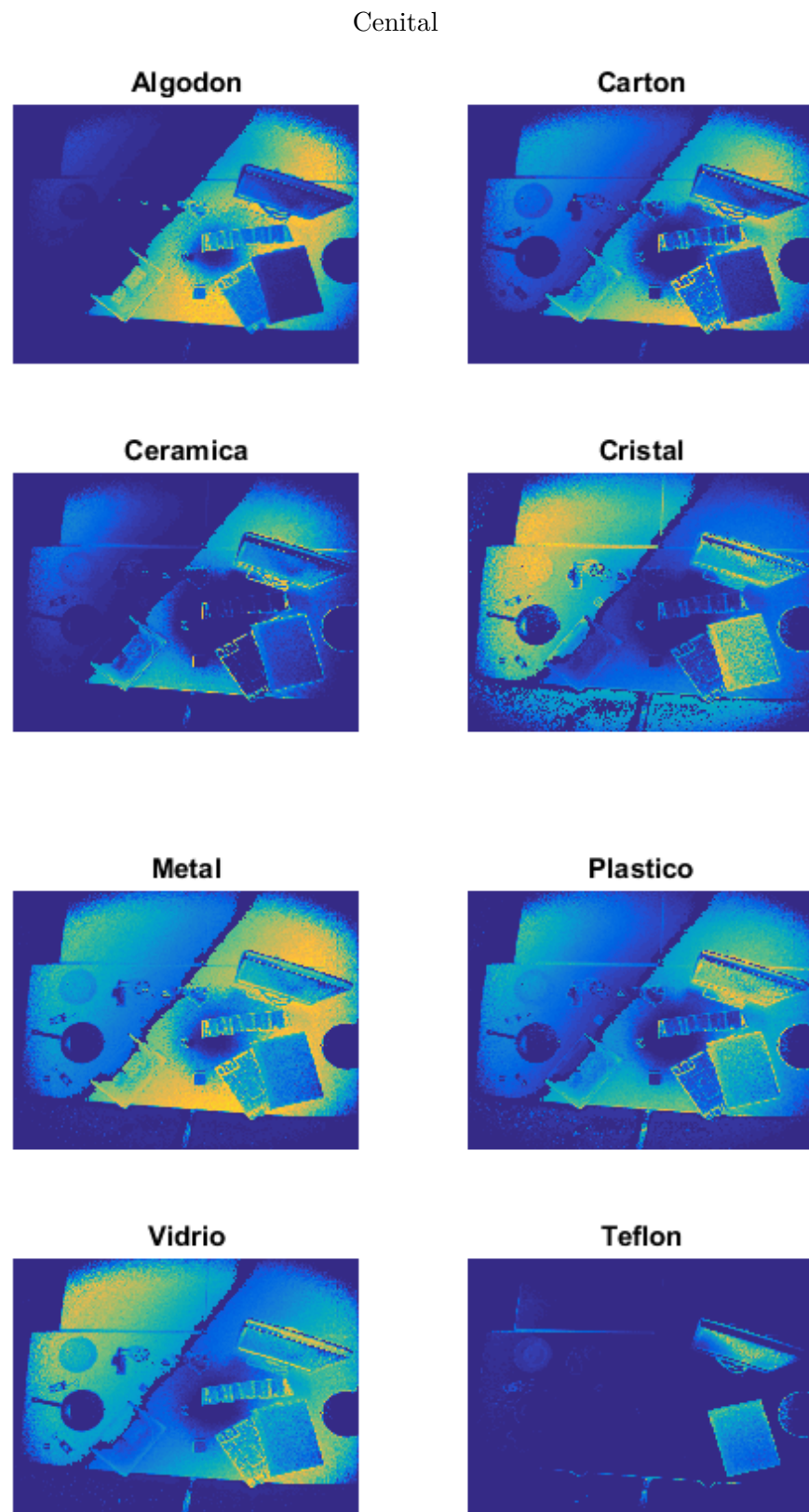


Figura 5.2: Resultado de haber probado la secuencia con vista cenital de evaluación sobre nuestros 8 modelos de objeto.

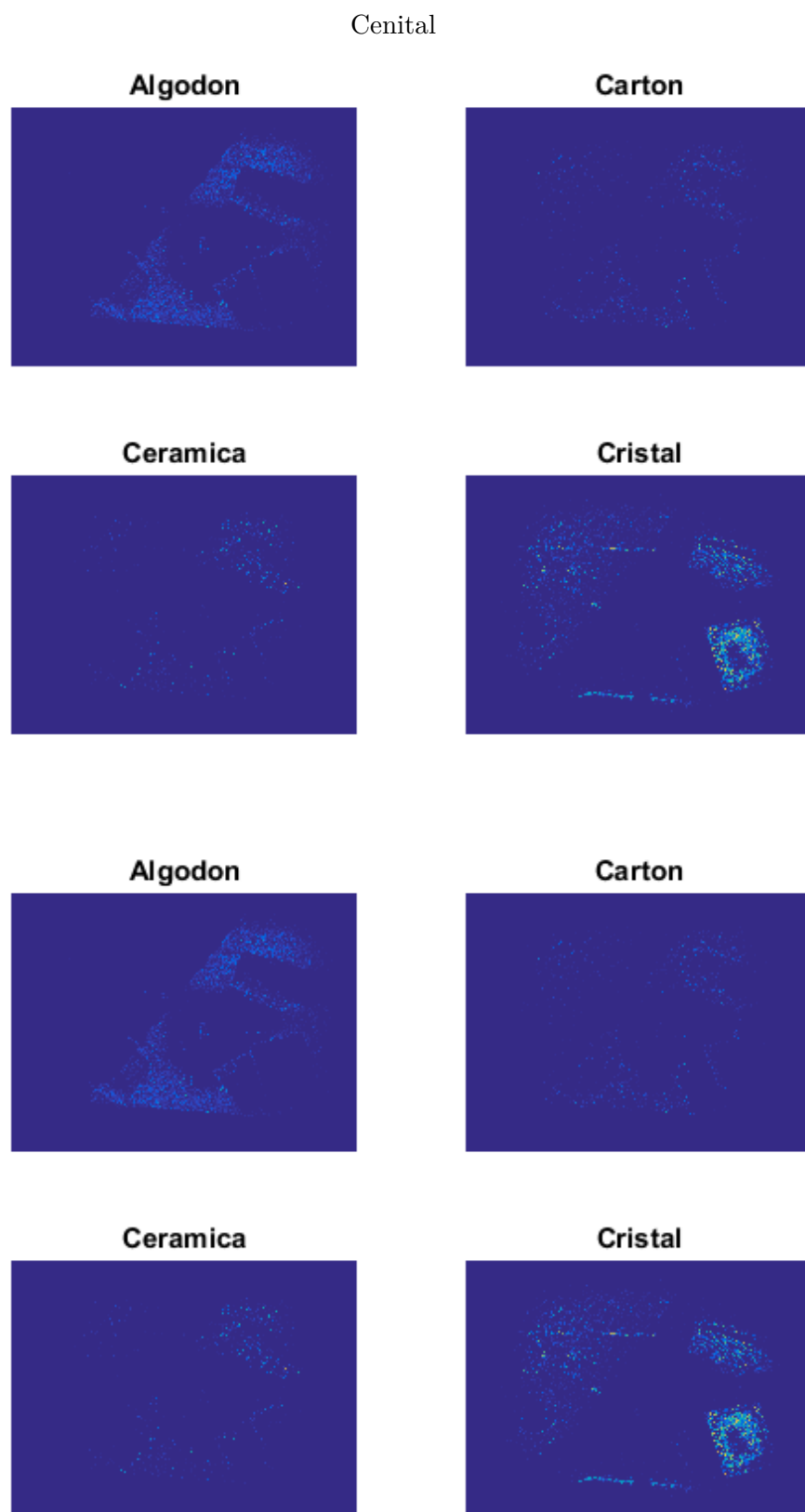


Figura 5.3: Resultado de haber probado la secuencia con vista cenital de evaluación sobre nuestros 8 modelos de objeto tras normalización.

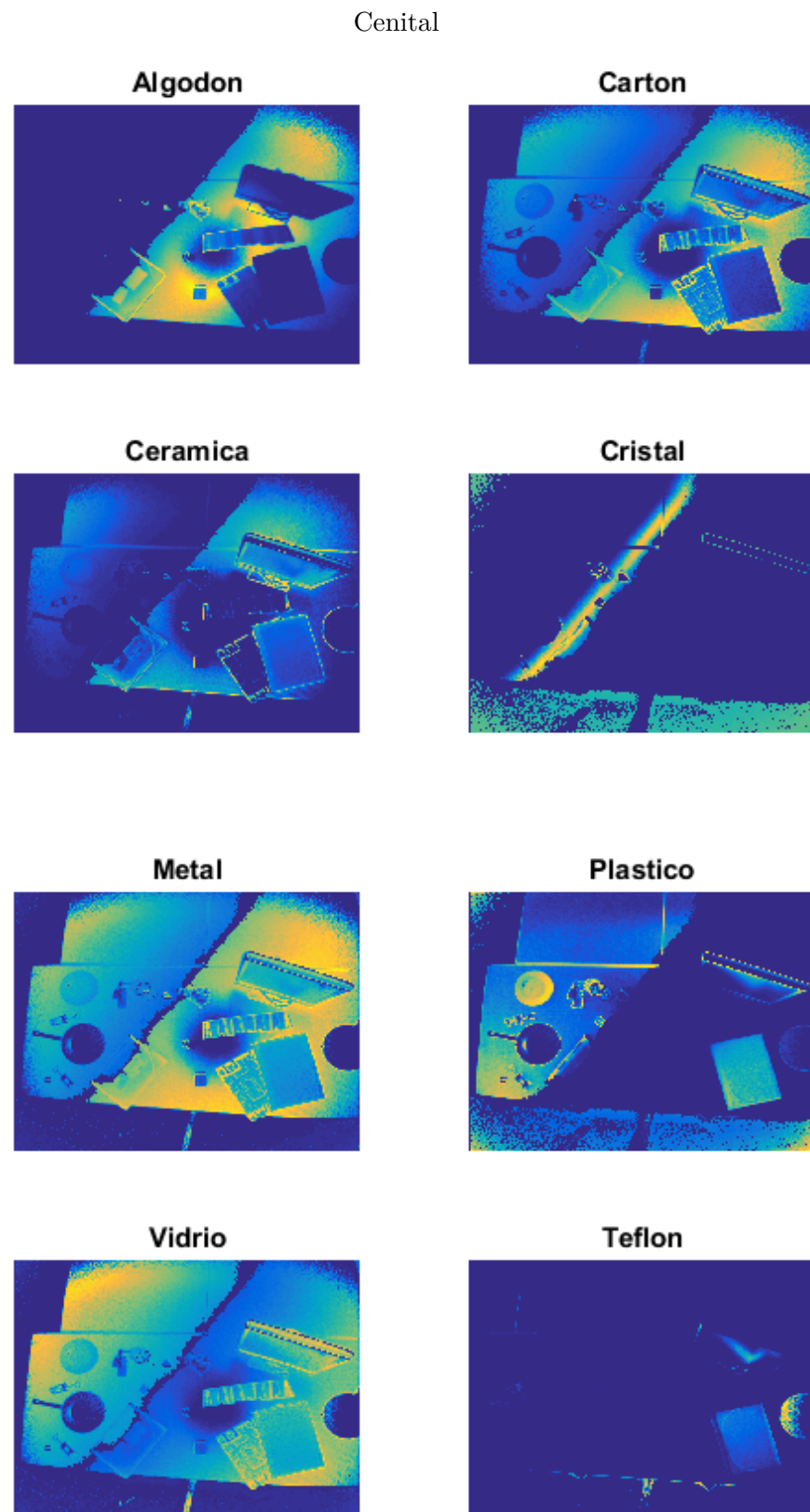


Figura 5.4: Resultado de haber probado la secuencia con vista cenital de evaluación sobre nuestros 8 modelos de objeto tras pca..

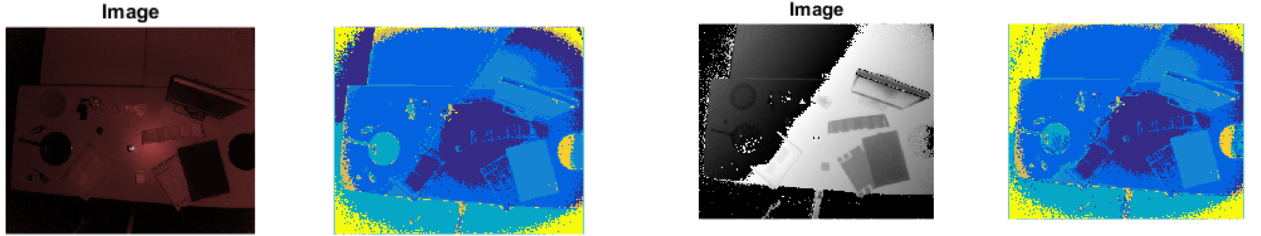


Figura 5.5: Media de las probabilidades de cada pixel mostradas como una sola imagen diferenciada en clases. (Columna izquierda infrarrojo y derecha profundidad). Etiquetas: Amarillo=Teflón, Azul=Cartón, Azul Oscuro=Algodón, Verde=Cristal, Naranja=Plástico.



Figura 5.6: Media de las probabilidades de cada pixel normalizados mostradas como una sola imagen diferenciada en clases. (Para infrarrojo y para profundidad). Etiquetas: Amarrillo y Verde=Cristal, Azul=Algodón.



Figura 5.7: Media de las probabilidades de cada pixel tras pca mostradas como una sola imagen diferenciada en clases. (Para infrarrojo y para profundidad). Etiquetas: Para infrarrojo: Azul=Algodón, Amarillo=Cristal | Para profundidad: Azul Oscuro=Algodón, Azul Claro=Cartón, Verde=Cristal, Naranja=Plástico, Amarillo=Teflón.

Capítulo 6

Conclusiones y trabajo futuro

6.1. Conclusiones

Se ha realizado un estudio del estado del arte, extenso y en profundidad para conocer como se ha realizado hasta ahora la detección de materiales y conocer como se puede aplicar a este trabajo de fin de grado en particular.

Además se ha generado una base de datos con imágenes de color, infrarrojo y profundidad de diversos materiales, desde diferentes puntos de vista y con cambios de iluminación, para posteriormente dividirla en una parte para entrenamiento y otra para evaluación. A continuación, se ha realizado un etiquetado de esta base de datos de forma manual para distinguir materiales en cada secuencia.

Con la información de los píxeles que pertenecen a cada material, se han obtenido descriptores mediante la transformada de *Fourier* que entrenan al modelo de conocimiento, de forma que tengamos un modelo por cada material.

Por último, se ha estudiado el funcionamiento de nuestro modelo, mediante el uso de los datos almacenados para evaluación de nuestro *Dataset*, y se ha realizado una discusión de los resultados.

6.2. Trabajo futuro

A la vista de los resultados, se propone profundizar en lo siguiente:

- Generación de un dataset más extenso, con más materiales y más situaciones.
- Incorporar la información de color a los modelos generados.
- Estudiar el funcionamiento de los descriptores para otros modelos diferentes a GMM.

Bibliografía

- [1] VVAA, “Microsoft kinect v2,” 2010. 5
- [2] VVAA, “Adaptador para pc de kinect v2,” 2010. 6
- [3] L. Sharan, R. Rosenholtz, and E. H. Adelson, “Accuracy and speed of material categorization in real-world images,” *Journal of Vision*, vol. 14, no. 10, 2014. 6
- [4] K. J. D. B. V. G. S. K. N. J. J. Koenderink, “Columbia-utrecht reflectance and texture database,” NA. 7
- [5] W. Li and M. Fritz, “Base de datos mpi-vips,” 2010. 8
- [6] J. Shlens, “A Tutorial on Principal Component Analysis,” *ArXiv e-prints*, Apr. 2014. 8
- [7] L. de la Fuente Crespo, “Análisis componentes principales,” NA. 9
- [8] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001. 9
- [9] J. R. Quinlan, “Induction of decision trees,” *Mach. Learn.*, vol. 1, pp. 81–106, Mar. 1986. 9
- [10] T. M. Mitchell, *Decision Tree Learning*. 1997. 9
- [11] C. wei Hsu, C. chung Chang, and C. jen Lin, “A practical guide to support vector classification,” 2010. 9
- [12] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, “An efficient k-means clustering algorithm: analysis and implementation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 881–892, Jul 2002. 9
- [13] L. Cayton and S. Dasgupta, “A learning framework for nearest neighbor search,” in *Advances in Neural Information Processing Systems 20* (J. C. Platt, D. Koller, Y. Singer, and S. T. Roweis, eds.), pp. 233–240, Curran Associates, Inc., 2008. 9
- [14] VVAA, “Clustering using gaussian mixture models,” NA. 9
- [15] D. K. Iakovidis, T. Goudas, C. Smailis, and I. Maglogiannis, “Ratsnake: A versatile image annotation tool with application to computer-aided diagnosis,” *The Scientific World Journal*, vol. 2014, Article ID 286856, 2014. 14

- [16] A. Davis*, K. L. Bouman*, J. G. Chen, M. Rubinstein, O. Bărbos, F. Durand, and W. T. Freeman, “Visual vibrometry: Estimating material properties from small motions in video,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, pp. 732–745, April 2017. [37](#), [38](#)
- [17] H. Zhang, K. Dana, and K. Nishino, “Reflectance hashing for material recognition,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3071–3080, June 2015. [39](#)
- [18] C. Liu, L. Sharan, E. H. Adelson, and R. Rosenholtz, “Exploring features in a bayesian framework for material recognition,” in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 239–246, June 2010. [39](#)

Apéndice A

Estudios publicados sobre detección de materiales

A continuación, se realiza un desarrollo de las investigaciones más relevantes que se han realizado hasta la fecha en identificación de materiales con aproximaciones similares a la que se realiza en este trabajo.

En el estudio publicado [16] se pretenden estimar las propiedades de un material a partir de pequeños movimientos en un video. Para ello, se realiza una grabación en video de telas y barras de diferentes materiales excitadas por pequeñas fuerzas como sonidos o corrientes de aire y, a continuación, se procede a extraer las señales de movimiento para crear un espectro del movimiento de cada objeto. En definitiva, este espectro contiene información que nos puede permitir extraer conclusiones de algunas propiedades de los objetos. En la siguiente imagen podemos observar la idea general del método. Figura:A.1

Para obtener estas características, el estudio hace uso de la teoría de vibración y más concretamente, del análisis de modos propios: Podemos modelar un objeto como un sistema de masas puntuales conectadas por muelles, lo cual se puede representar con la siguiente ecuación diferencial. EcuaciónA.1:

$$M\ddot{x} + Kx = 0 \quad (\text{A.1})$$

Donde M es una matriz de masas, \ddot{x} será la aceleración, K es una matriz de la rigidez de los muelles y x representa el desplazamiento. Buscaremos soluciones de esta ecuación de la forma. Ecuación A.2:

$$x = A\sin(\omega t + \varphi) \quad (\text{A.2})$$

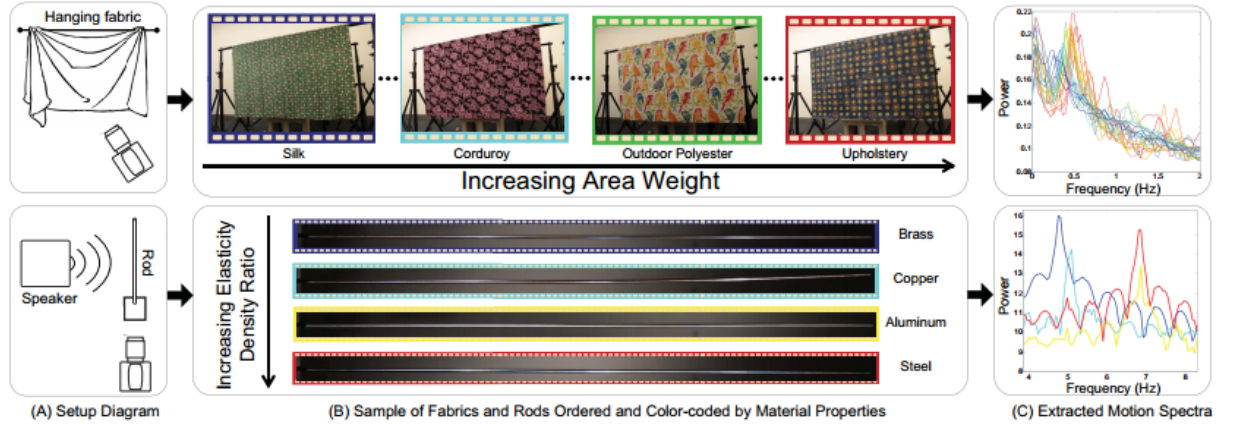


Figura A.1: Idea general del método.(Extraído de [16])

Obteniendo la siguiente solución. Ecuación A.3

$$[K - \omega_i^2 M] A_i = 0 \quad (A.3)$$

Donde ω_i son las frecuencias naturales de nuestro objeto (autovalores) y A_i nos describe los modos de vibración que ocurren a esas frecuencias (autovectores). Si cambiamos la forma del objeto, cambiará la dispersión de M y K por lo que también cambiarán los autovalores y los autovectores. Pero si mantenemos la forma del objeto y tan solo cambiamos las propiedades, las matrices M y K tan solo sufrirán un escalado lo que implica que los autovalores cambiarán, pero los autovectores seguirán siendo los mismos.

Es decir, diferentes objetos con la misma geometría tendrán los mismos modos propios pero sus frecuencias de resonancia serán diferentes dependiendo de las propiedades del material. Por lo que la idea será aprender las propiedades de los materiales a partir de los valores de dichas frecuencias ω_i .

A continuación, para la extracción de propiedades de los materiales se utilizarán dos estrategias:

1. Para las barras se usará un esquema de votación ya que es un acercamiento muy preciso, pero requiere que se conozca la estructura observada (la simple geometría de una barra sujeta hace fácil solucionar los modos de vibración como una función de longitud, diámetro, densidad y módulo elástico)
2. Para las telas, se entrenará un modelo con datos de entrenamiento que posteriormente serán evaluados mediante los datos de test.

Como podemos observar, el uso de componentes frecuenciales para la extracción de características de los materiales es una idea ya utilizada, aunque nunca se ha llegado a realizar con la aportación de información de profundidad e infrarrojo.

En [17] tenemos un estudio detallado sobre el uso del patrón de reflectancia de los materiales para su caracterización. En este artículo, se confirma que en la tarea de reconocimiento de materiales no es suficiente utilizar tan solo información sobre el color y la geometría de un material, sino que es necesario hacer uso de otras informaciones.

Para ello, se utiliza una cámara de reflectancia basada en un diseño que hace uso de un espejo parabólico a la salida, esta cámara devuelve un disco de reflectancia: un denso muestreo de la reflectancia superficial del material proyectado en una sola imagen.

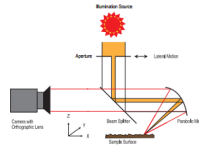


Figura A.2: Cámara de reflectancia basada en espejo parabólico (Extraído de [17])

A la vista de los resultados empíricos, se observa que con el uso del patrón de reflectancia se puede realizar una buena identificación de materiales incluso para conjuntos pequeños de entrenamiento. Este resultado tiene implicaciones importantes para aplicaciones de identificación en tiempo real, ya que este acercamiento puede soportar entrenamiento online para conjuntos de entrenamiento compactos.

En [18] vemos un estudio de identificación de materiales sobre la base de datos “Flickr Material Database” basándose en el uso de aLDA (Augmented Latent Dirichlet Allocation). Las características extraídas se cuantifican como palabras visuales y cada imagen se trata como una bolsa de palabras y se utiliza LDA para modelar la distribución de palabras visuales para cada categoría. En la figura se puede observar el funcionamiento completo del algoritmo: Figura: A.3

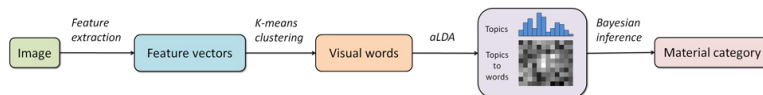


Figura A.3: Funcionamiento del aLDA completo. (Extraído de [18])

Apéndice B

Resto de vistas para infrarrojo y profundidad.

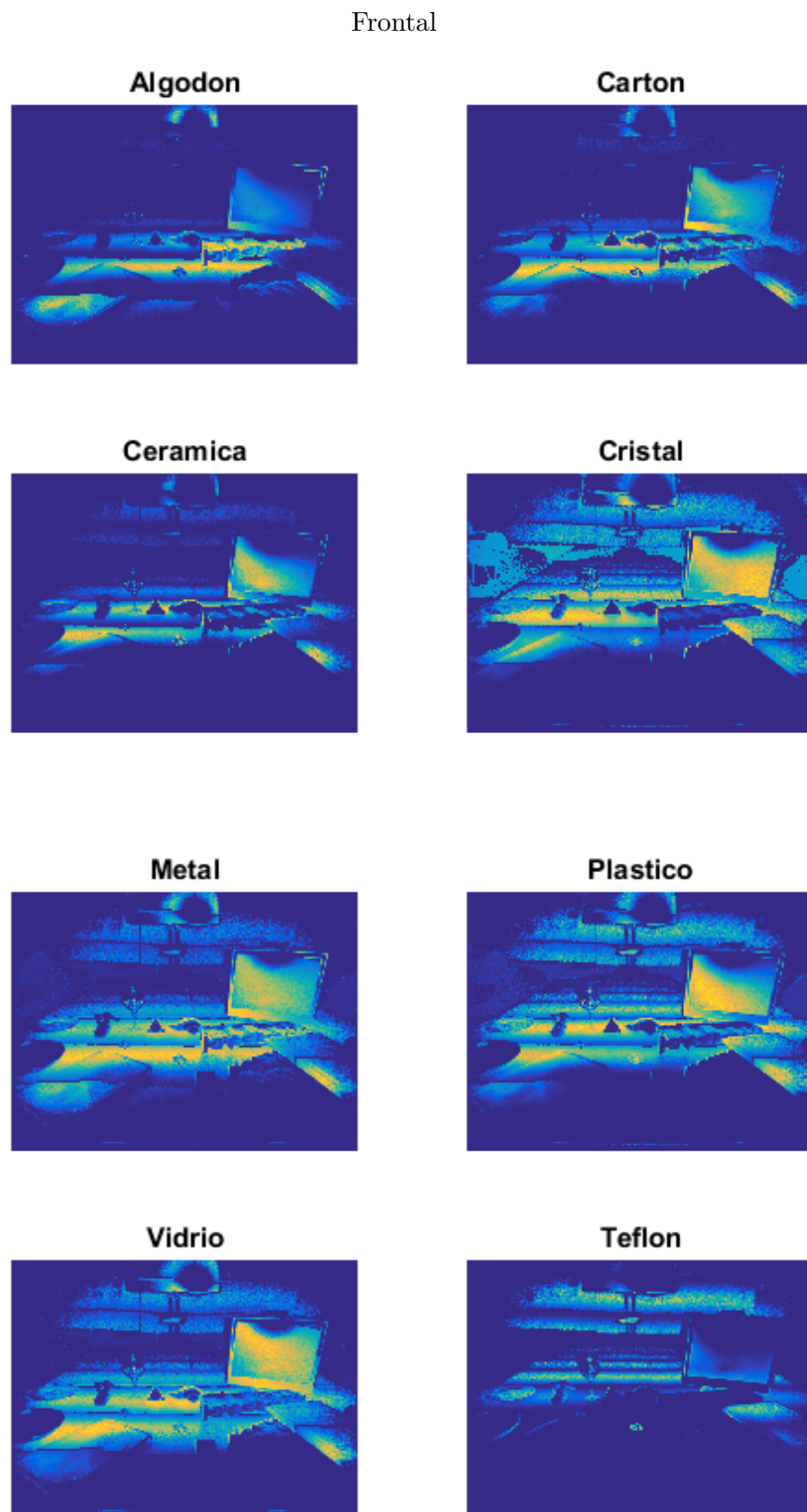


Figura B.1: Resultado de haber probado la secuencia con vista frontal de evaluación sobre nuestros 8 modelos de objeto

Lateral

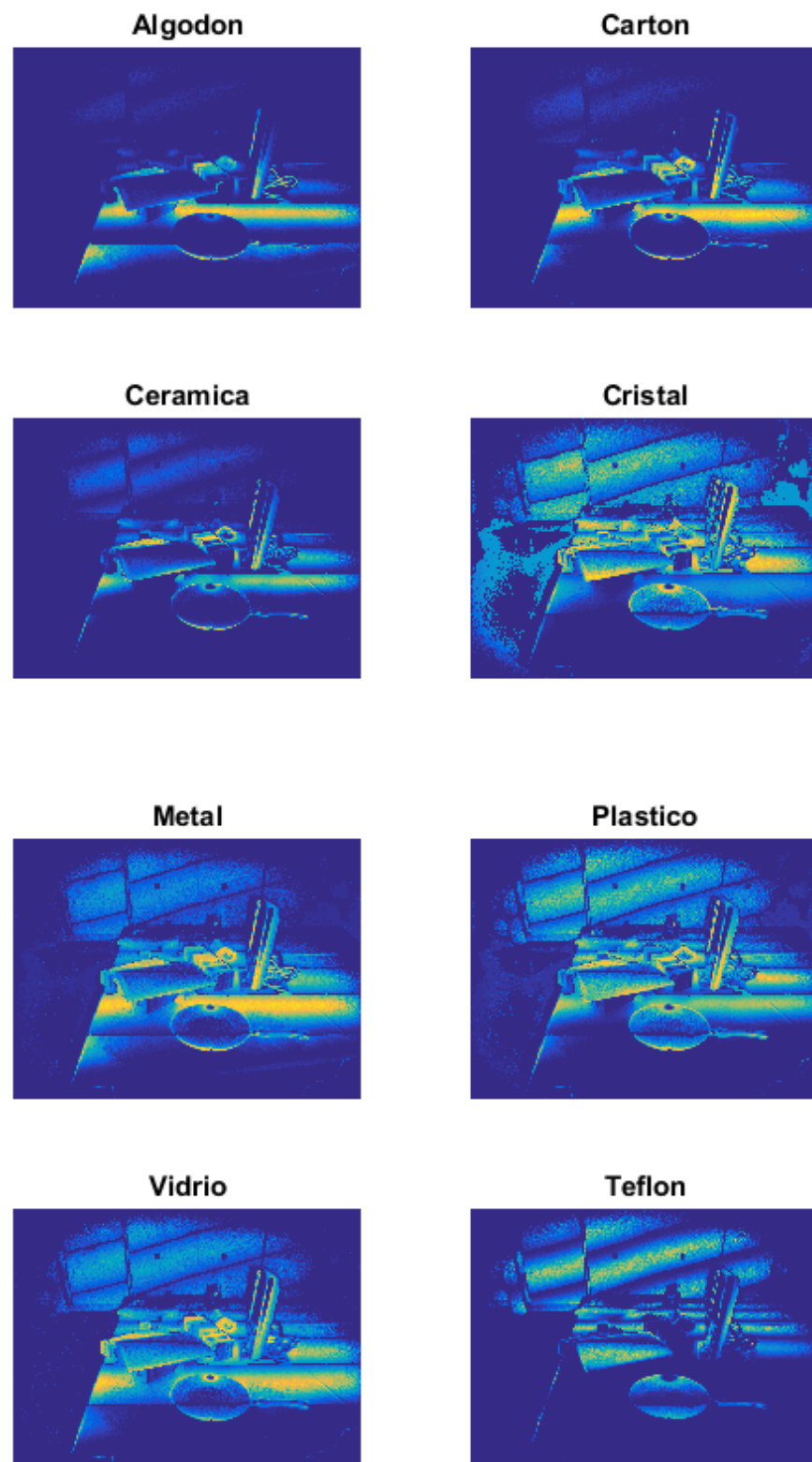


Figura B.2: Resultado de haber probado la secuencia con vista cenital de evaluación sobre nuestros 8 modelos de objeto

Pseudolateral

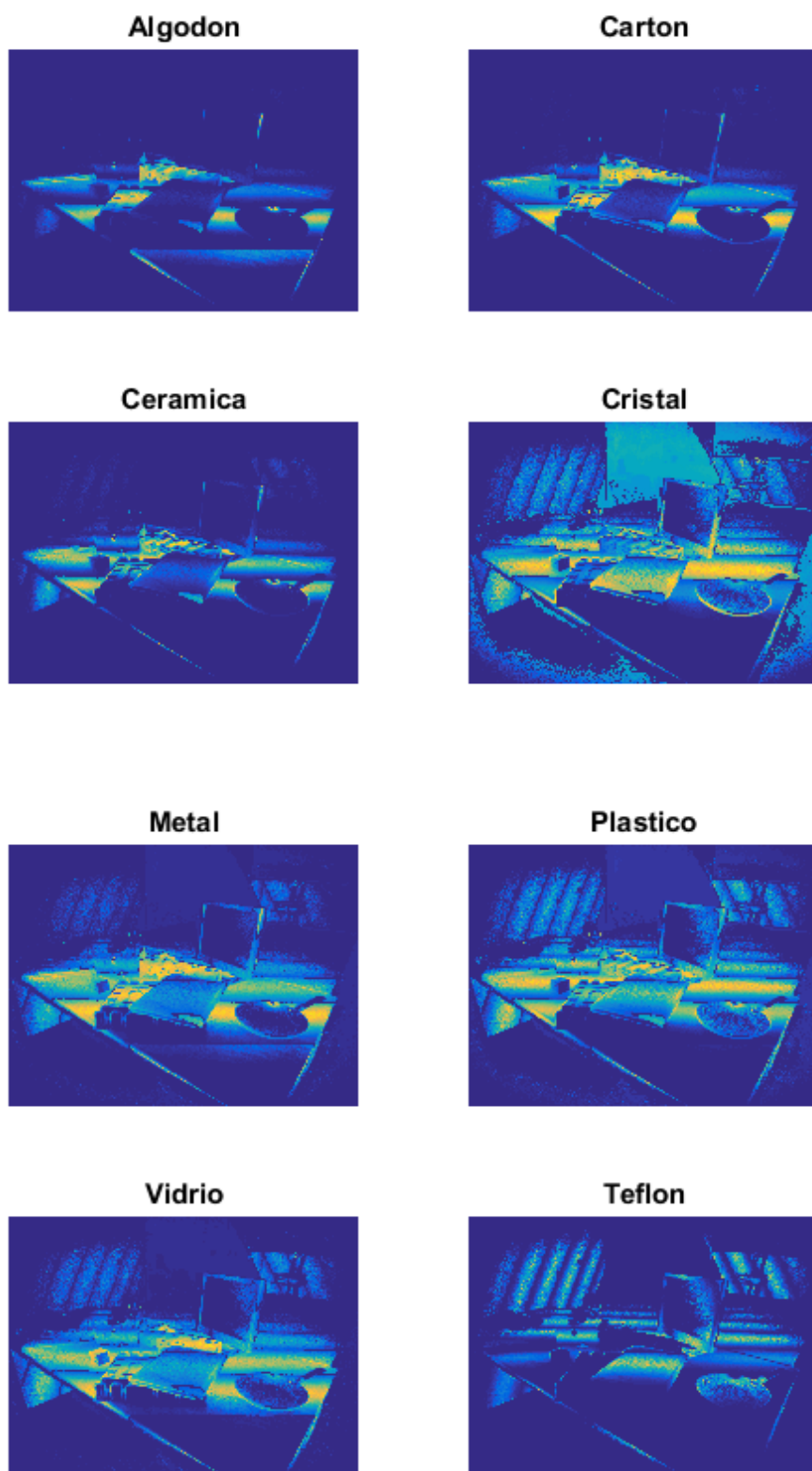


Figura B.3: Resultado de haber probado la secuencia con vista pseudolateral de evaluación sobre nuestros 8 modelos de objeto

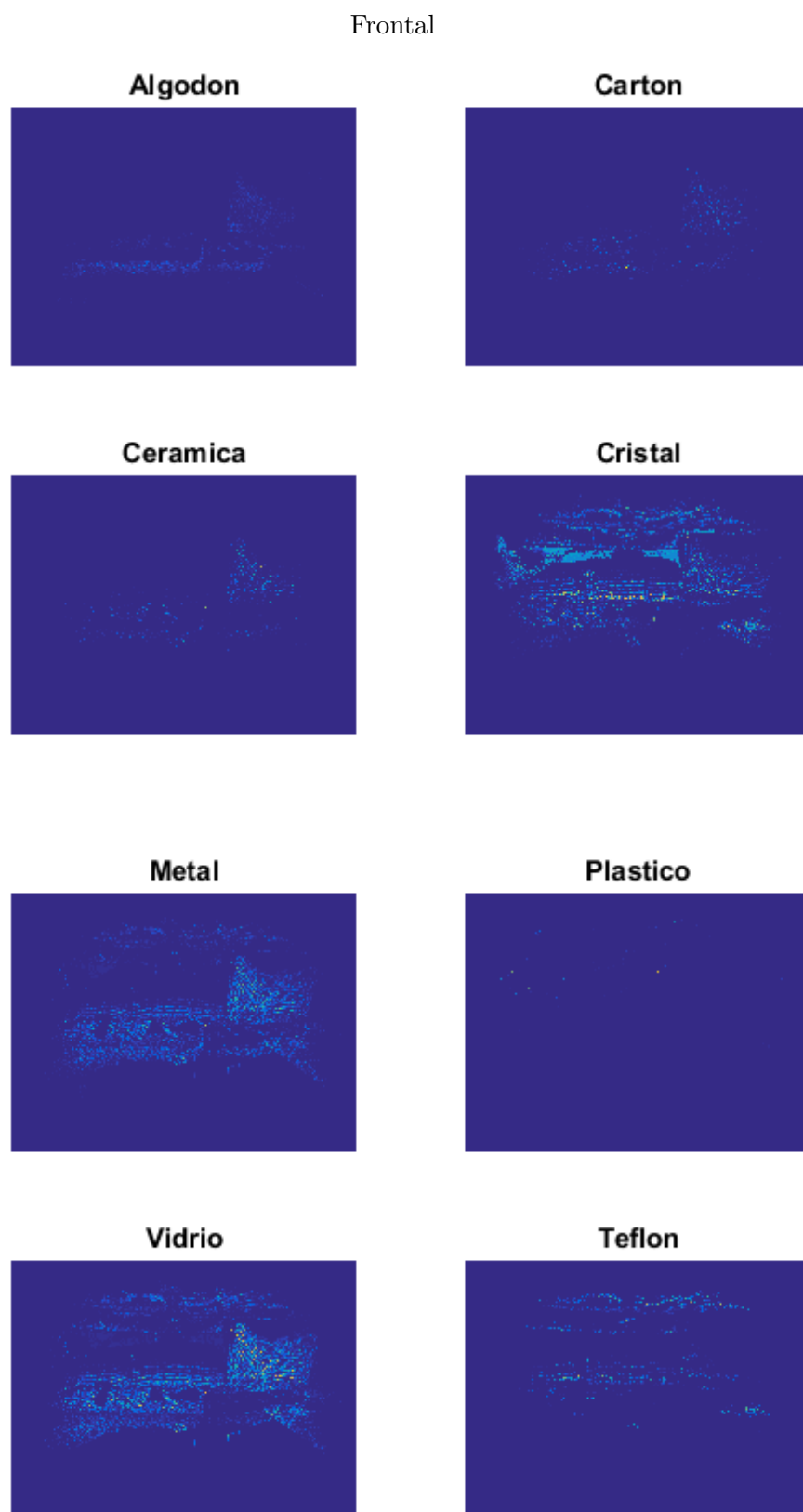


Figura B.4: Resultado de haber probado la secuencia con vista frontal de evaluación sobre nuestros 8 modelos de objeto tras normalización

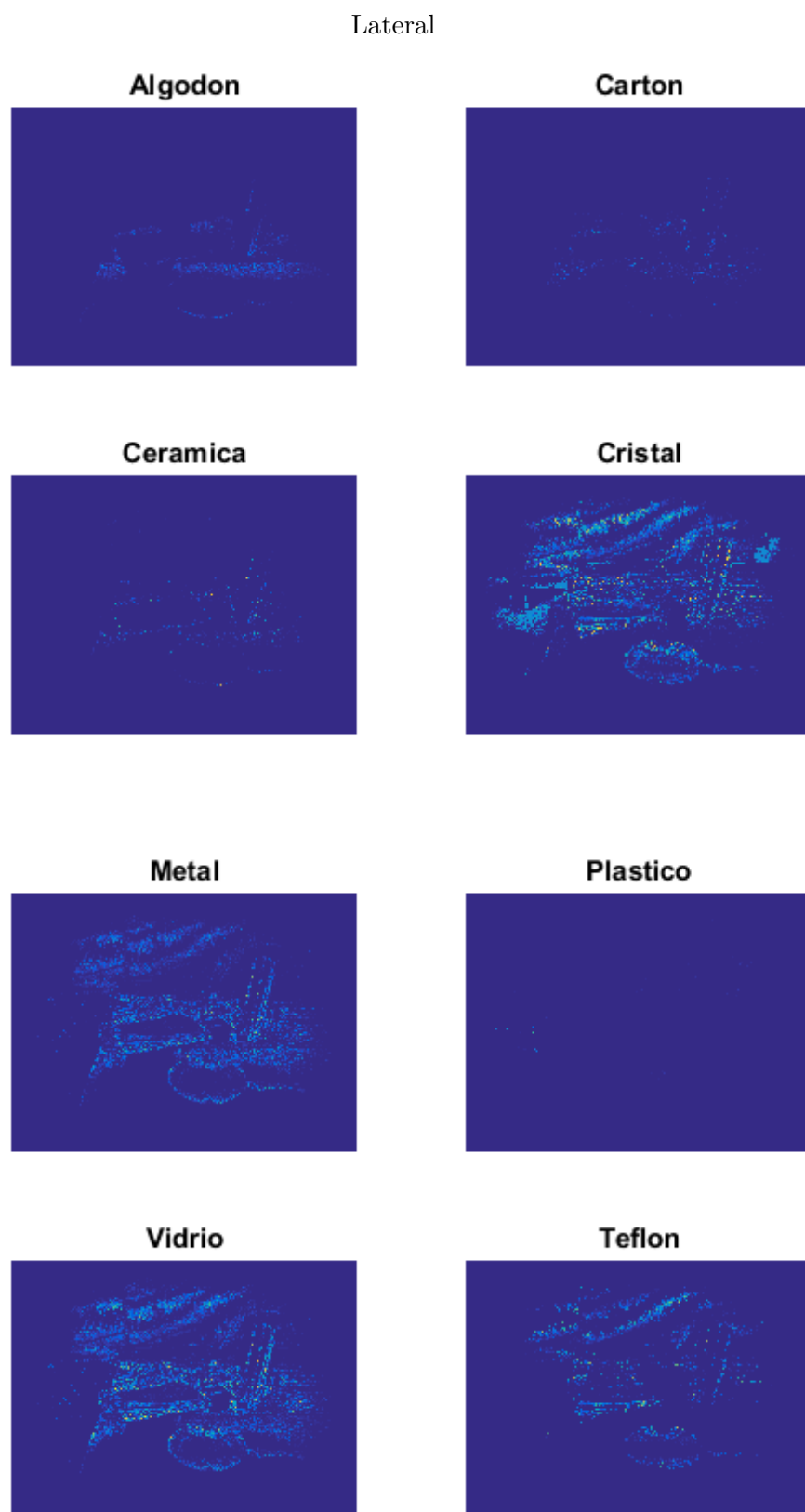


Figura B.5: Resultado de haber probado la secuencia con vista cenital de evaluación sobre nuestros 8 modelos de objeto tras normalización

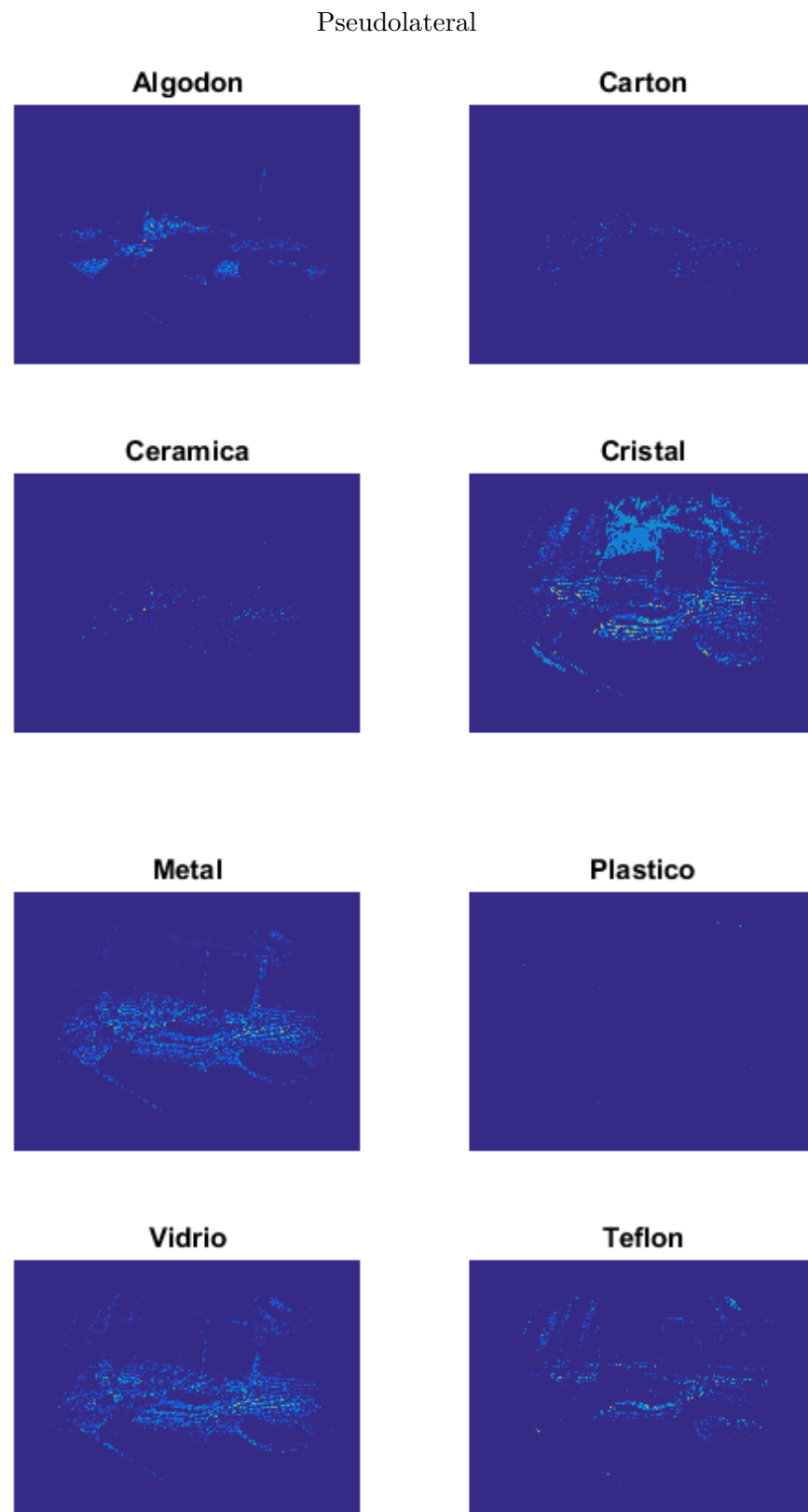


Figura B.6: Resultado de haber probado la secuencia con vista pseudolateral de evaluación sobre nuestros 8 modelos de objeto tras normalización.

Frontal

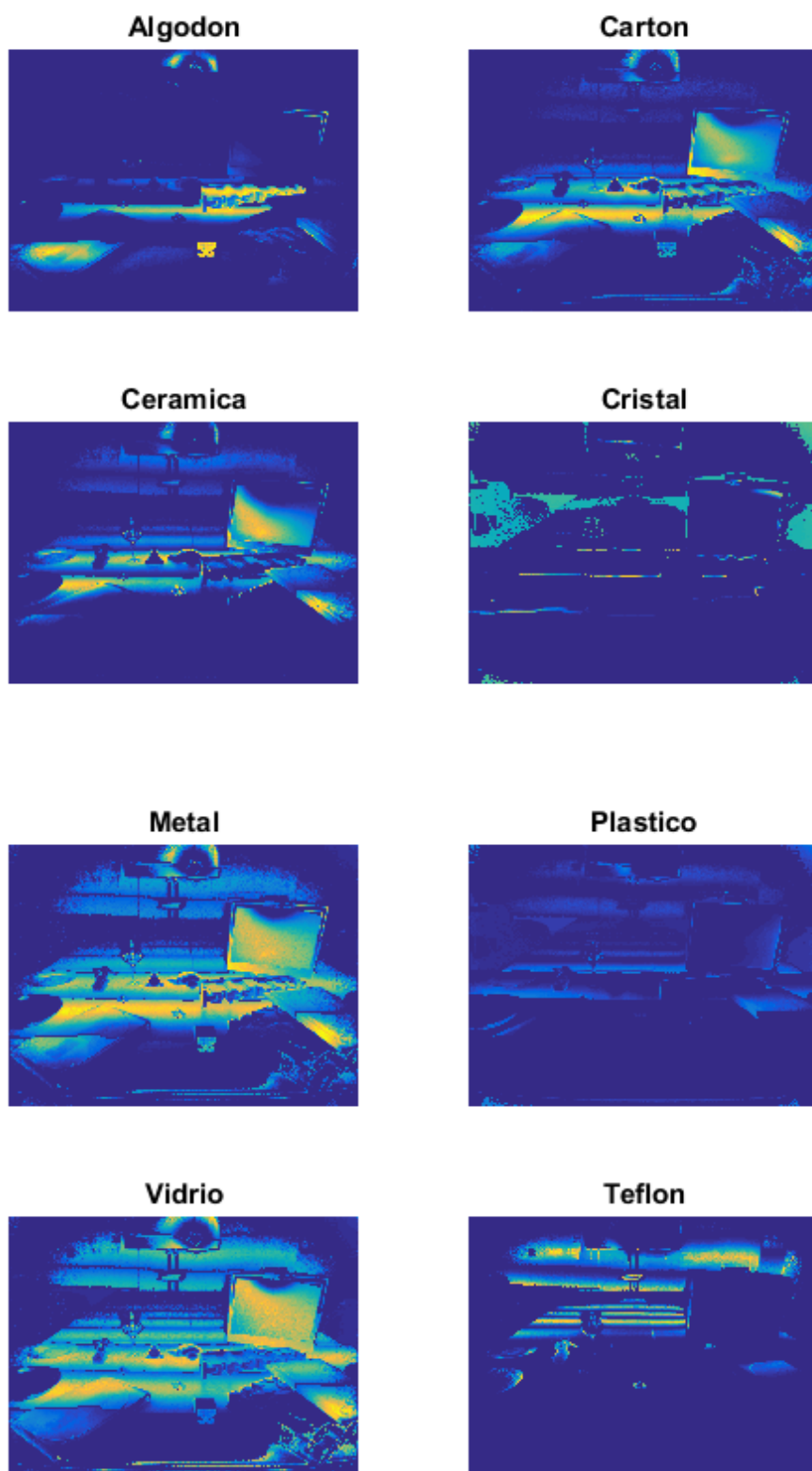


Figura B.7: Resultado de haber probado la secuencia con vista frontal de evaluación sobre nuestros 8 modelos de objeto tras pca.

Lateral

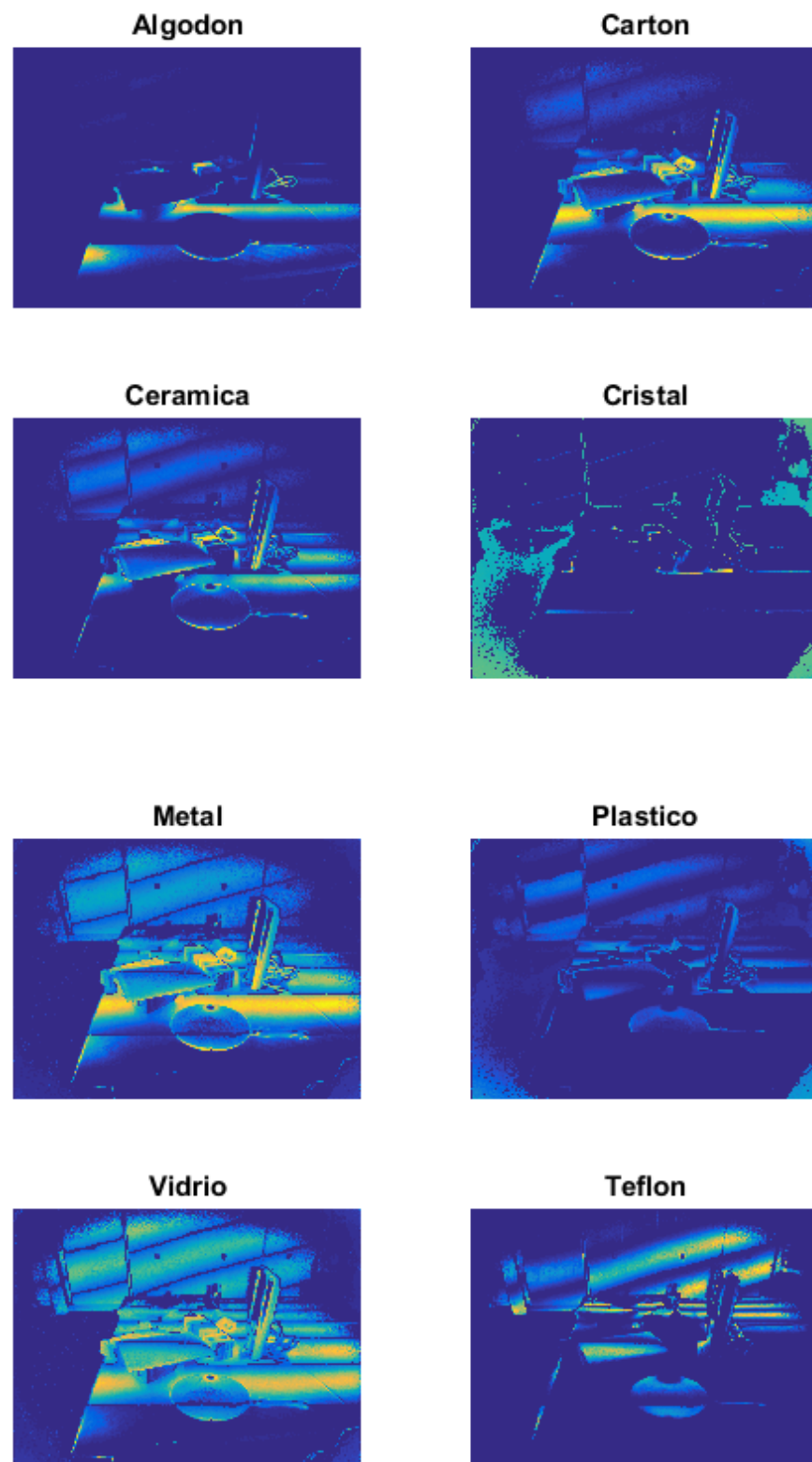
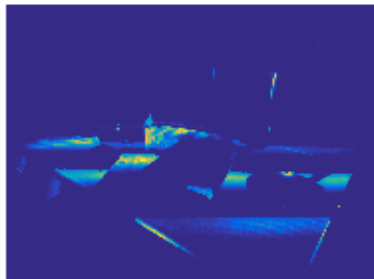


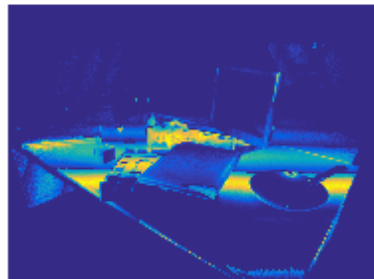
Figura B.8: Resultado de haber probado la secuencia con vista cenital de evaluación sobre nuestros 8 modelos de objeto tras pca

Pseudolateral

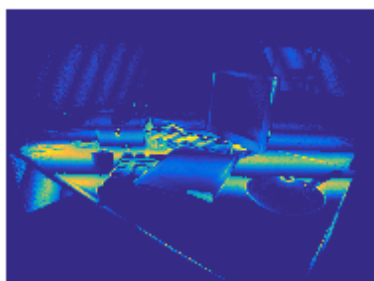
Algodon



Carton



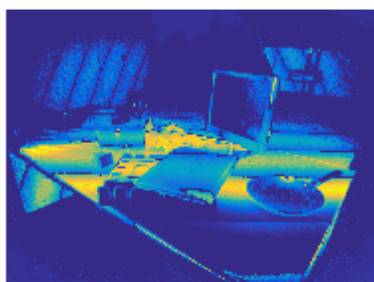
Ceramica



Cristal



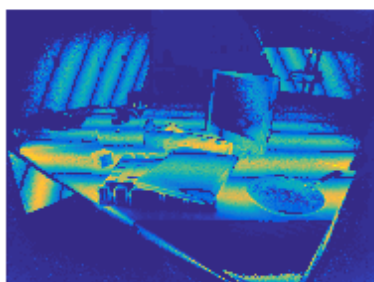
Metal



Plastico



Vidrio



Teflon



Figura B.9: Resultado de haber probado la secuencia con vista pseudolateral de evaluación sobre nuestros 8 modelos de objeto tras pca.